RELATIONSHIPS BETWEEN PEER RATINGS AND QUANTITATIVE PROPERTIES OF DEFENSE IN-HOUSE LABORATORIES

VOLUME I

by

John H. Walker, Jr.

Management Analysis Report 71-1
Office for Laboratory Management
Office of the Director of Defense Research and Engineering
Washington, D. C. 20301

FOREWORD

This report represents the culmination of an assignment that began in March of 1970, when the author was detailed from the Naval Weapons Laboratory, Dahlgren, Virginia, to the Office of the Director of Defense Research and Engineering (Laboratory Management). The assignment was, simply stated, to determine if there were meaningful relationships between peer ratings - which had already been obtained - and quantitative properties of laboratories - for which data existed for fiscal years 1967, 1968, and 1969.

The purpose of the report is to describe the computation of the peer ratings and to examine the rankings upon which they are based; to describe the elements of the laboratory resources data base and to examine their distribution among the various laboratories; and to describe the investigation of relationships between the peer ratings and the quantitative properties. In this latter respect, the report has a two-fold purpose: to inform people of the findings, and to serve as a basis for further study.

The report has been published in two volumes. Volume I contains a narrative description of the different phases of the study. Volume II contains various tables and other data which were generated in the course of the work; these have been presented as appendices.

Preceding page blank

ACKNOWLEDGEMENTS

Many people have contributed to the development and preparation of this report. I am grateful to the management of the Naval Weapons Laboratory for providing the opportunity to work on the project, and especially for bearing with me through the preparation of the report. I wish to thank Ed Glass and Evan Anderson of the Office of Laboratory Management for their help and encouragement throughout the assignment. I also wish to thank Bert Levy and John Marsh of the Harry Diamond Laboratories for their consultation and assistance, John McCloskey of the University of Dayton Research Institute for comments and suggestions, and Rich Hein and Frank Reynolds of the Air Force Information Systems Division for programming support.

At the Naval Weapons Laboratory, I am indebted to Marlin Thomas and Gary Gemmill for their help with the statistical analyses (and for trying to straighten out my statistical warp); to Lottie Anderson, Dave Wolper, and Marvin Oliver for their patience in the face of many program changes and re-runs; to Judy Gautier for similar feats of patience during the preparation of the manuscript; and to Goldie Hayden, Retha Bullock, and the staffs of the Computer Operations Branch and the Printing Branch.

In particular, I wish to thank Ray Brancolini, who was involved in almost all phases of the project, for his timely and ingenious help with the programming and computation; Dorothy Elam, who programmed the composition of the report, and made many helpful editorial suggestions; and my wife, Margaret, who provided the necessary moral support, and helped with many different phases of the work.

Preceding page blank

CONTENTS

Preceding page blank

EXECUTIVE SUMMARY

Introduction

The effective management of R&D laboratories requires a continual appraisal of what they are doing and what they should be doing. Within the Department of Defense, many different types of appraisals are regularly made - supervisory evaluations, program evaluation, special appraisals, committee visits, etc. Most of these techniques are subjective in nature and lack a quantitative basis, particularly for comparisons among laboratories with widely differing missions and technical orientation. To rectify some of the deficiencies in these appraisal systems, the laboratory resources data base was developed to provide comparative statistical and trend data on the characteristics and performance of laboratories. It was felt that the utility and significance of these data might be improved if they could be related to the comparative technical competence or quality of laboratories.

For this purpose, professional technical people with a substantial degree of industrial, university, or Federal laboratory experience - mostly in the management of R&D programs and organizations - were asked to rank the laboratories according to their opinion of a laboratory's ability to perform its assigned mission. Emphasis was placed on the technical rather than the administrative background of the rankers so that in their judgement of a particular organization, consideration would be given more to technical competence than to administrative efficiency. Because the background and experience of the participants were generally comparable to those of the managers of the laboratories being ranked, the rankings have been called "peer rankings", or more commonly, "peer ratings".

It was not the intention of the survey to develop a precise rank ordering, but rather to obtain a measure of relative laboratory quality which might be used in the exploration of relationships between technical reputation and measurable characteristics of laboratories. A recognition and awareness of such relationships, where meaningful, can assist laboratory managers in formulating relevant policies and practices appropriate to their particular environments The purpose of this report is to describe in more detail how the peer rankings were obtained, to show how they were subsequently used to obtain a relative ratings for each laboratory, and to summarize various studies conducted using the peer rankings and the quantitative laboratory properties.

Preceding page blank

The Peer Ratings

The peer ratings were obtained during fiscal year 1969 by means of a survey conducted by Edward M. Glass and Evan D. Anderson, following a procedure used by Maurice A. Apstein in 1963 (reference [1]). The three hundred and twenty-five peers - program managers, laboratory directors, commanding officers, technical specialists, and scientists and engineers - ranked those of the seventy-nine DoD laboratories with which they were familiar. The composition of the raters was as follows:

OSD (mainly DDR&E)	44
Service Headquarters and Commands	64
Laboratories	136
Private Industry	66
Other Sources	15
	325

The present study was limited to an examination of the fifty-one physical sciences and engineering laboratories. There was a relatively strong association between the ratings of Navy laboratories and the *number* of times they were ranked, a mild association for the Army laboratories, and none for the Air Force laboratories (probably because they had the least variation in the number of rankings).

Proportionally, the rankings by the OSD participants were equally distributed among the three military departments. The Army and Navy raters tended to rank mainly the laboratories within their respective departments. The Air Force participants ranked more laboratories outside the Air Force than in the Air Force, but on a proportional basis, they ranked considerably more of their own than did the raters from the other two services. The participants from the private sector also tended to rate proportionally more of the Air Force laboratories.

The laboratories ranked in the first five were clearly among those considered "best"; the laboratories ranked among the last eight were generally rated low by most of the rater groups. Where comparisons were possible, the ratings agreed fairly well with a similar survey made by Maurice Apstein in 1963. (The coefficient of rank-order correlation between the Apstein survey and the present one was .95 for twenty-six laboratories.)

The Laboratory Properties

The quantitative properties of laboratories indicate the number of personnel. the number of professionals. the number with advanced degrees. etc. They tell the size of the plant. cost of equipment. and value of property. They spell out the different types of appropriations. their source, and whether they are in-house or out-of-house research or development. etc. Also included are data on the number of patent applications. papers published. meetings attended, technical reports. and graduate training.

For the present study. the data used were from fiscal years 1967. 1968, and 1969. Most of the analyses were based on the 1968 data. or upon the average of the three years. For the most part, the personnel data are relatively stable and change rather slowly; the financial data are subject to much greater annual variation.

The appropriations data generally reflect the different funding practices among the three military departments. as is indicated by the following percent distribution of appropriations among the majority of the physical sciences and engineering laboratories in each service. averaged over fiscal years 1967. 1968. and 1969:

	Army	Navy	Air Force
Research and Development	82	71	97
Procurement	10	13	-
Operations and Maintenance	4	7	-
Miscellaneous	4	9	3

Other differences. based on the data for fiscal year 1968, are reflected in the percentages of in-house to out-of-house activity, as well as the percentage of RDT&E in research and exploratory development:

	Army	Navy	Air Force
Total In-House Dollars	47	60	23
Total Out-of-House Dollars	53	40	77
In-House RDT&E	50	70	23
Out-of-House RDT&E	50	30	77
T6.1/RDT&E	7	9	16
T6.2/RDT&E	25	29	56

Relationships Between Peer Ratings and Laboratory Properties

Using the data for fiscal year 1968, it was found that there are significant correlations between the peer ratings and the properties of the Navy laboratories. The Army laboratories for the most part show only minor correlations between the peer ratings and the laboratory properties. A few of the Air Force correlations are as large as those found for the Navy, but because of the much smaller number of Air Force laboratories, most of the correlations are not as statistically significant.

By examination of the correlations between the peer ratings and the other years of the data base, it was found that in many cases they are quite similar for each of the three years of the data base, and also for the averages of the data over the three year period. The correlations with the data for fiscal year 1968 generally lie between those for fiscal year 1967 and fiscal year 1969, and thus fiscal year 1968 appears to have been a fortuitous choice of a base year.

A study was also made of the differences in correlations between the laboratory properties and the ratings based on the rankings of different rater groups. It was found that there was considerable variation among the various groups, particularly between the raters from industry and the raters from DoD. The correlations based upon the industrial ratings are for the most part lower than those based on the DoD ratings, but the effect on the overall ratings is not as large as it might seem, since the DoD group comprises more than 75% of the sample. On the average, the larger of the correlations based on ratings from all the rater groups are within 5% of the corresponding DoD correlations.

Because the distribution of properties among the various laboratories was asymmetrical - in all cases, no more than five laboratories accounted for more than twenty-five percent of each property - a number of experiments were conducted to determine the effect of unusually large values of the variables. These experiments, using the data for fiscal year 1969, showed that a fairly large number of the correlations changed substantially when outliers were removed. Overall, however, there were relatively few cases among the Army or Navy elements where an extreme point exerted undue force in (1) raising a correlation to a significantly high value, or (2) masking out significant correlations in the remaining variables. The marginal number of Air Force laboratories precludes making a similar statement, one way or the other, about the effect of extrema on their correlations.

Except for a few instances, the Army correlations were not overly dependent on either the largest or the highest-ranked laboratories; the Navy correlations showed a slight dependency on both. The Air Force laboratories were generally

higher without the largest laboratory, and lower without the highest-ranked laboratory. Some experiments with a subset of the Army laboratories indicate that in selected circumstances the correlations may be comparable to those found for the Navy variables.

The correlation of the peer ratings with the logarithms of the laboratory properties were generally somewhat smaller than the correlations with the untransformed variables, although the distribution was more centralized and less asymmetrical. This suggests that the correlations are dependent upon the size of the laboratory properties.

Dividing the properties by various normalizing factors, e.g., the number of professionals, the size of the R&D program, etc., substantially reduced the correlation of the Navy variables - again indicating a dependency on size - but tended to raise the correlations with the Air Force variables. Some of the more significant of the Air Force correlations appear to result from the large proportion of research and exploratory development dollars in their R&D appropriations (more than 70%. compared to about 35% each for the other two military departments), and from the relatively large ratio of out-of-house R&D to in-house R&D in the Air Force laboratories (more than 3:1, compared to 1:1 and 3:7 for the Army and the Navy).

An examination of the quantitative properties for fiscal year 1968 for each military department showed that the highest-ranked laboratories generally had more than double the amount of the lowest-ranked laboratories. For several variables, such as the number of PhD's and the amount of research dollars, the few top-rated DoD laboratories had six to ten times as much of the property as the few bottom-rated laboratories.

When viewed on a per-professional basis, the ratios for these same variables were about half as large - the highest-rated laboratories having from three to five times the proportion of the lowest-rated ones. There was a large variation in the proportions of the total RDT&E program between the high- and low-rated laboratories of the different services. The Army ratios were about the same: 69.1 thousand dollars per professional in the upper laboratories, 68.7 thousand dollars per professionals in the lower ones. The corresponding Navy and Air Force numbers were 64.8 : 39.8 and 86.4 : 119.0, respectively.

In the majority of the data elements for each military department, with the laboratories ordered according to their proportion of the element (or property), there are many instances where a high-rated laboratory is adjacent to a low-rated laboratory; thus any statements of the sort "the high-rated laboratories have this

much, whereas the low-rated laboratories have that much" must be taken advisedly. Generally this has been provided for in this report by phrases such as "the higher-rated laboratories tend to ...".

Another way to examine the data is to observe the number of times the low-ranked laboratories appear in the high-order positions when the properties are ranked by size, and vice versa. The variables showing the least number of such occurrences, i.e., the variables such that high corresponds to high and low corresponds to low, are Total Research Appropriations, Funding From Non-DoD Sources, the Number of Professionals with Masters Degrees, and the combined size of the Research and Exploratory Development Appropriations. When the data are normalized by dividing by the total number of professionals, the variables that most consistently are in the proper position (by the above criteria) are Research Dollars, Research and Exploratory Development Appropriations, Equipment, and Scientific Equipment Acquisition.

Regression analyses, using the individual rankings, show that a linear regression equation can account for about 30% of the variation in the Navy ratings, about 20% of the variation in the Army ratings, and about 12% of the variation in the Air Force ratings. Alternatively, if one uses the mean values of the rankings disregard the statistical variability between raters, then one can account for 90% of the variations with eight Army variables, three Navy variables, and three Air Force variables. However, these can be selected in a variety of ways, and will generally not be independent of one another, so that the use of the regression equation to "control" the quality of the laboratories is quite unlikely. Yet in a very real sense, the candidate variables are all representative of control variables, for they are measures of the basic laboratory resources. Hence the inputs, to the regression equation, and consequently the outputs - and correspondingly, the peer ratings at some future date - are in some way a function of these basic elements.

Conclusions

It seems obvious that the technical competence of a laboratory depends much more upon the quality of its leadership, the vitality of its mission, and the enthusiasm and capability of its people, than upon the number of people or the size of its technical program. However, the numbers of people and the amounts of dollars, plant, and equipment are the laboratories' basic resources; and in the aggregate, the professional quality of the staff and the nature of the technical program are reflected by the proportion of personnel with advanced degrees and the proportion of funding for research and development. For the Navy laboratories, there are substantial correlations (of the order .800 and higher) between the peer rating and laboratory elements such as Professionals with Advanced Degrees,

Equipment, Scientific Equipment Acquisition, and the In-House RDT&E Program. The correlations are based more on the size of the particular properties than on their generic proportions. The particular properties are so intimately associated with R&D capability that it must be concluded that the ratings are meaningfully related to the R&D competence of the Navy laboratories.

For the Army and the Air Force laboratories, the association between peer ratings and quantitative properties is not clear. The Army laboratories consist of a larger number of extreme variables; several of the higher-rated laboratories have quite different characteristics. Nine of the Army laboratories have less than two hundred professionals, which makes discrimination between them difficult. A few correlations are somewhat higher when normalized by the number of professionals, but even then they are only marginally useful. The Air Force laboratories, on the other hand, are more similar in the distribution of their characteristics, but their ratings are also more closely bunched than those of the other two services, tending to be more in the upper middle part of the distribution of the ratings. Also, the relatively small number of laboratories reduces the statistical significance of correlations which are of the same magnitude as those of the other two services, and which are therefore *seemingly* as meaningful.

There is some suggestion, when considering the few top-ranked and bottom-ranked laboratories on an overall DoD basis, that the raters may have put a premium on the research aspects of laboratory activity, partiuclarly with respect to the number of PhD's and the magnitude of the research program. It is also possible that this is a secondary rather than a primary effect, i.e., the laboratories having the larger research appropriations may also be the most widely known, and are being cited by renown (this is not to argue that the quality of the research program was not initially responsible for the renown).

The reputations of the laboratories change slowly with time; all but three of the twenty-nine laboratories for which there were corresponding ratings from the Apstein 1963 survey were in remarkably good agreement with the ratings of the present survey. This spot-lights one of the unanswered questions of the present study: which of the years of the data base typifies the raters' knowlege of the laboratories? The ratings were coincident with fiscal year 1969; but even assuming that the raters had current knowledge of the laboratories' technical competence, is the state of that competence dependent upon current values of the resources, or does it reflect the resources that were available two, five or ten years earlier? More realistically, the raters may have had current knowledge of only a portion of the laboratories they rated, making even more uncertain the lag between resources potential and laboratory accomplishment.

A linear regression equation might conceivably be used for predicting the future rating of the laboratories, but it would be necessary to conduct one or two more surveys, for calibration and validation, before one could hope to arrive at a meaningful regression equation. Further, it would be necessary to determine a function to represent the time lag between a rater's estimation of a laboratory's technical quality and its actual present capability. As to the value of such a model if it existed, it might be a useful mangement tool for answering "what if" types of questions; but on the whole, I doubt that it would have practical utility.

Recommendations

A peer ranking survey similar to the one described herein should be conducted within a three to five year period of the 1969 survey. The participants should identify themselves as before, and additionally according to service affiliation. Care should be taken to ensure that the different rater groups are given appropriate representation. A follow-up interview should be conducted with a sub-sample of the participants in order to obtain insight to the various alternatives that were considered during the ranking process.

The laboratory resources data base should continue to be maintained and expanded according to the needs of its users. With the addition of the data for fiscal year 1971, the laboratory properties data base will span a period of five fiscal years, and should be able to provide comprehensive data upon which to base analysis and prediction of past, present, and future trends.

PART I

Introduction to the Peer Ratings and the Quantitative Laboratory Properties

1. INTRODUCTION

1.1 Nature and Purpose of Study

During the summer and fall of 1969, the Office of Laboratory Management conducted a survey to determine the comparative technical competence of the seventy-nine Department of Defense (DoD) laboratories. The survey was conducted by Evan D. Anderson, following a procedure used by Maurice A. Apstein [1] in a similar survey in 1963. Professional technical people with a substantial degree of industrial, university, or Federal laboratory experience - mostly in the management of R&D programs and organizations - were asked to rank the laboratories according to their opinion of a laboratory's ability to perform its assigned mission. Emphasis was placed on the technical rather than the administrative background of the rankers so that in their judgement of a particular organization, consideration would be given more to technical competence than to administrative efficiency.

The rankings obtained from each participant were divided into deciles, and these were then used to compute the laboratory ratings. Because the background and experience of the participants were generally comparable to those of the managers of the laboratories being ranked, the rankings have been called "peer rankings", or more commonly, because of the underlying methodology and the subsequent transformation to ratings, "peer ratings".

It was not the intention of the survey to develop a precise rank ordering, but rather to obtain a measure of relative laboratory quality which might be used in the exploration of relationships between technical reputation and measurable characteristics of laboratories. A recognition and awareness of such relationships, where meaningful, can assist laboratory managers in formulating relevant policies and practices appropriate to their particular environments.

The study described in this report attempts to look for meaningful relationships between the peer ratings and quantitative properties of laboratories such as staffing, funding, property, equipment, etc. The methodology of the survey and highlights of the study have been reported by Edward M. Glass in references [2] and [3]. The purpose of this report is to describe in more detail how the peer rankings were obtained, to show how they were subsequently used to obtain a relative rating for each laboratory, and to summarize various studies conducted using the peer rankings and the quantitative laboratory properties.

The remaining sections of the introduction describe the conduct of the survey, touch briefly upon salient characteristics of the DoD laboratories, and review the

chronology of the study. Chapter 2 covers the computation of the peer ratings; Chapter 3 elaborates upon the elements of the data base (the quantitative laboratory properties): and Chapters 4 through 8 describe the analyses conducted to date; Chapter 9 contains comments, summary, conclusions, and recommendations.

1.2 Conduct of the Survey

Some five hundred people were invited to participate in the peer rating survey; three hundred and twenty-five responded. The majority of these were from within the Department of Defense, although a substantial minority (about 25%) were from universities, industry, and other government laboratories. The participants were initially categorized into seven groups:

(1) Program managers and technical specialists in the Office of the Secretary of Defense (principally in the Office of the Director of Defense Research and Engineering (DDR&E))
(2) Program managers and technical specialists on the headquarters staffs of the military departments
(3) Program managers and technical specialists on the staff of the various military commands
(4) DoD laboratory managers (mainly Technical Directors and Commanding Officers)
(5) R&D managers and technical specialists outside the Department of Defense but within the Federal Government
(6) Technical specialists, R&D managers, consultants, and professionals from private industry and from nonprofit organizations
(7) Scientists and engineers in academic institutions

Each participant received an instruction sheet (Figure 1.1) and a deck of cards containing the names of the laboratories (Figure 1.2). Where convenient, the forms were presented in person; otherwise, they were mailed to the selected participants. The identities of the individual participants were not recorded, although information was obtained as to which of the seven types of activity was represented. Those in categories (2) and (3) were further identified as to military department, as were some but not all of those from category (4).

The participants were asked to first separate the cards into two piles, according to whether or not they knew enough about a laboratory to give it a rank. (To "know" a laboratory was defined as being sufficiently acquainted with its work to have formed an opinion regarding the technical competence of the entire laboratory to perform its assigned mission or the technical competence of any segment to accomplish its assigned mission).

INSTRUCTIONS FOR PEER RATING OF DoD LABS

1. Here is a list of the major R&D installations in the Department of Defense. Please separate them into two piles; those you know and those you do not. For purposes of this exercise, to "know" a laboratory is defined as being sufficiently acquainted with its work to have formed an opinion regarding the technical competence of the entire laboratory to perform its assigned mission or the technical competence of any segment to accomplish its assigned mission. This opinion need not have been obtained first hand, it may have been formed through reading government reports, technical articles in the open literature, and via inputs from other scientific professionals whose judgment you respect. (If there is any question in your mind regarding the validity of your information, place the card in the "unknown" pile.)

2. Discard the "unknown" pile and separate the known pile into three groups, ABOVE AVERAGE, AVERAGE, BELOW AVERAGE.

3. Now take the ABOVE AVERAGE pile and lay them out in front of you so that they are all in view. Place them in ranking order by selecting first, the BEST of the group, then the next best and so on until you have ranked the entire group. Place this pile aside.

4. Now take the AVERAGE pile and separate it into two groups; ABOVE AVERAGE and BELOW AVERAGE. Then rank each group as in 3.

5. Repeat with the BELOW AVERAGE group.

6. Combine all piles in ranking order, and consecutively number the computer cards to indicate the ranked position of each laboratory in the upper right hand corner.

7. In the upper left hand corner of the No. 1 card indicate the type of organization you are with. For example if you are in a laboratory just insert "LAB". If you are in a headquarters activity insert the appropriate name or symbol such as AMC, CNM, AFSC, Army Staff, Navy Staff, Air Staff, ORA, ARO, ONR, private industry, etc.

8. Please do not mutilate the cards.

FIGURE 1.1
INSTRUCTION SHEET

DoD Laboratories

DAD001	LIMITED WAR LAB., ABERDEEN P.G., MD.
DAD002	BEHAVIORAL SCIENCE RES. LAB., WASH., D.C.
DAD004	ENGINEER TOPOGRAPHIC LAB., FT BELVOIR, VA.
DAD005	ENGINEER WATERWAYS EXPERIMENT STA., VICKSBURG, MISS.
DAD008	AEROMEDICAL RESEARCH UNIT, FT. RUCKER, ALA.
DAD009	DENTAL RES. INST., WALTER REED AMC, WASH., D.C.
DAD010	MEDICAL BIOMECHANICAL RES. LAB., W. REED AMC, WASH DC
DAD011	MED. RES. AND NUTR. LAB., FITZSIMONS GH, DENVER, COLO.
DAD012	MEDICAL RES. UNIT, EUROPE (LANDSTUHL, GERMANY)
DAD013	MEDICAL EQUIP. R+D LAB., FT. TOTTEN, FLUSHING, N.Y.
DAD014	MEDICAL RES. LAB. FT. KNOX, KY.
DAD015	RES. INST. OF ENVIRONMENTAL MEDICINE, NATICK, MASS.
DAD016	WALTER REED ARMY INST. OF RES., WASH., D.C.
DAD019	MEDICAL UNIT, FT. DETRICK, MD.
DAD020	MEDICAL RES. UNIT, PRESIDIO OF SAN FRANCISCO, CAL.
DAD021	AERO. RES. LAB., AMES RES. CTR., MOFFETT FIELD, CAL.
DAD022	BALLISTIC RES. LABS., ABERDEEN PROVING GROUND, MD.
DAD023	COATING AND CHEMICAL LAB., ABERDEEN PROVING GROUND, MD.
DAD024	TERRESTRIAL SCIENCES CENTER, HANOVER, N.H.
DAD025	HARRY DIAMOND LABS., WASH., D.C.
DAD026	HUMAN ENG. LABS., ABERDEEN PROVING GROUND, MD.
DAD027	MATERIALS AND MECHANICS RESEARCH CENTER, WATERTOWN, MASS.
DAD028	NUCLEAR DEFENSE LAB., EDGEWOOD ARSENAL, MD.
DAD029	NATICK LABS., NATICK, MASS
DAD030	AVIATION MATERIEL LABS., FT. EUSTIS, VA.
DAD032	ELECTRONICS LABS., FT. MONMOUTH, N.J.
DAD033	MISSILE COMMAND LABS, REDSTONE ARSENAL, ALABAMA
DAD034	MOBILITY EQUIP. RES. AND DEVELOP. CENTER, FT BELVOIR, VA.
DAD035	FRANKFORD ARSENAL LABS., PHILA., PA.
DAD036	PICATINNY ARSENAL LABS., DOVER, N.J.
DAD037	BIOLOGICAL LABS., FT. DETRICK, MD.
DAD038	EDGEWOOD ARSENAL LABS., EDGEWOOD ARSENAL, MD
DAD039	ROCK ISLAND ARSENAL LABS., ROCK ISLAND, ILL
DAD040	WATERVLIET ARSENAL LABS., WATERVLIET, N.Y.
DAD041	TANK-AUTOMOTIVE COMMAND LABS., WARREN, MICH.
N00122	NAVAL UNDWTR. WEAPONS RES. AND ENG. STA., NEWPORT, R.I.
N00156	NAVAL AIR ENGINEERING CENTER, PHILADELPHIA, PA.
N00167	NAVAL SHIP RESEARCH AND DEVELOPMENT CTR., WASH., D.C.
N00173	NAVAL RESEARCH LABORATORY, WASHINGTON, D.C.
N00178	NAVAL WEAPONS LABORATORY, DAHLGREN, VA.
N00953	NAVAL ELECTRONICS LABORATORY CENTER, SAN DIEGO, CALIF.
N0751A	NAVAL AEROSPACE MEDICAL INST., NAMC, PENSACOLA, FLA.
N60530	NAVAL WEAPONS CENTER, CHINA LAKE, CALIF.
N60921	NAVAL ORDNANCE LABORATORY, WHITE OAK, MD.
N61331	NAVY MINE DEFENSE LAB, PANAMA CITY, FLORIDA
N61339	NAVAL TRAINING DEVICES CENTER, ORLANDO, FLA.
N61533	NSRDC(ANNAPOLIS) - (MARINE ENGINEERING LAB)
N61T51	NAVAL MEDICAL RESEARCH UNIT NO.3, CAIRO, EGYPT
N61776	NAVAL SUBMARINE MEDICAL CTR., NEW LONDON, GROTON, CONN
N62145	NAVAL MEDICAL FIELD RESEARCH LAB., CAMP LEJEUNE, N.C.
N62269	NAVAL AIR DEVELOPMENT CENTER, JOHNSVILLE, PA.
N62374	NAVAL MEDICAL RESEARCH UNIT NO.4, GREAT LAKES, ILL
N62399	NAVAL CIVIL ENGINEERING LABORATORY, PORT HUENEME, CAL.
N62462	NAVAL APPLIED SCIENCE LABORATORY, BROOKLYN, N.Y.
N62479	NAVAL RADIOLOGICAL DEFENSE LAB., SAN FRANCISCO, CALIF.
N62718	NAVAL WEAPONS CENTER CORONA LAB, CORONA, CAL.
N62814	NAVAL MEDICAL RESEARCH UNIT NO.2, TAIPEI, TAIWAN
N63116	NAVY MED. NEUROPSYCHIATRIC RES. UNIT, SAN DIEGO, CAL.
N63190	NAVAL UNDERSEA WARFARE CENTER, PASADENA, CALIF.
N63379	NAVAL PERSONNEL RESEARCH ACTIVITY, SAN DIEGO, CALIF.
N64223	NAVAL MEDICAL RESEARCH INSTITUTE, NNMC, BETHESDA
N66127	NAVAL PERSONNEL PROGRAM SUPPORT ACTIVITY, WASH., D. C.
N70024	NAVY UNDERWATER SOUND LABORATORY, NEW LONDON, CONN.
OFC105	MATERIALS LAB., WPAFB, OHIO
OF0121	AEROSPACE RESEARCH LABORATORIES, WPAFB, OHIO
OFC123	FRANK J. SEILER RESEARCH LAB., ACADEMY, COLO.
OF0124	CAMBRIDGE RESEARCH LABS., L.G. HANSCOM FLD., MASS.
OF0210	6571 AEROMEDICAL LAB., HOLLOMAN AFB, N.M.
OF0212	SCHOOL OF AVIATION MEDICINE, BROOKS AFB, TEX.
OF0213	6570 AEROSPACE MEDICAL RES. LAB., WPAFB, OHIO
OF0226	AEROSPACE MEDICAL LAB. (CLINICAL), LACKLAND AFB, TEX
OF0301	ROME AIR DEVELOPMENT CENTER,GRIFFIS AFB, N.Y.
OF0303	AVIONICS LAB., WPAFB, OHIO
OF0427	HUMAN RESOURCES LAB., BROOKS AFB, TEX.
OF0506	WEAPONS LAB., KIRTLAND AFB, N.M.
OF0602	AERO PROPULSION LAB., WPAFB, OHIO
OF0604	FLIGHT DYNAMICS LAB., WPAFB, OHIO
OF0606	ROCKET PROPULSION LAB., EDWARDS AFB, CALIF.
OF0607	ARMAMENT LAB., EGLIN AFB, FLA.

FIGURE 1.2

Next the participants were asked to discard the unknown pile and to separate the known pile into three groups: Above Average, Average, and Below Average. Additionally, since Apstein had observed that the middle group tended to have a larger number of members than the other two, the participants were asked to further divide the "Average" group into upper and lower sections. Then starting with the Above Average pile, the participants were asked to place the laboratories in ranking order by selecting first, the best of the group, then the next best, and so on. After completing this process for all piles, the participants were asked to combine them in ranking order and to consecutively number the cards starting with the first. The laboratory ratings were then computed by taking the average of the rankings received by each laboratory, linearly distributed on a scale from ten to zero for each rater.

It has been claimed that the survey technique measures technical reputation rather than technical competence. Apstein's assumption was that technical reputation in the scientific community is based upon the quality of scientific work. Therefore, the two terms were considered to be synonymous. However, depending upon how literally the participants followed the instructions, it is possible that a laboratory may be doing a competent job with respect to its mission, but if the principal elements of that mission are not esteemed by the population of raters, the laboratory may be ranked at the low end of the rating scale.

1.3 The DoD Laboratories

The Research, Development, Test and Evaluation program of the Department of Defense amounted to 7.8 billion dollars in 1967, to 7.9 billion dollars in 1968, and to 7.8 billion dollars in 1969. About one-third of this effort was conducted through the Department's one hundred and thirty in-house RDT&E activities.[1] These in turn performed slightly over one-half of the work at their own facilities, contracting the balance to other in-house facilities, to other federal activies, and to universities, private industry, etc.

Various data describing the facilities, programs, staffing, and funding of the 130 DoD RDT&E activities are collected annually as described in DODI 7700.9 (Appendix A). These data are summarized in an annual publication entitled "Department of Defense In-House RDT&E Activities" (reference [4]). A description of the data elements and an examination of their characteristics is presented in Chapter 3.

[1] The number of activities varies slightly from year to year, according to the rate at which existing installations are closed or consolidated and/or new ones are created.

The 130 activities are categorized into two parts - R&D laboratories and test activities. The study described in this report focuses primarily on the seventy-nine R&D laboratories. In fiscal year 1969 the seventy-nine R&D laboratories had real property and equipment valued at 2.3 billion dollars, employed 69,000 military and civilian personnel, and had a total annual program of 2.6 billion dollars. They employed three-quarters of the in-house professionals (including 92% of those with degrees at the doctorate level), and accounted for more than two-thirds of the RDT&E program (see Figure 1.3). In particular, more than 95% of their RDT&E effort was in research, exploratory development, advanced development, and engineering development (6.1, 6.2, 6.3, and 6.4) programs.

SUMMARY

DEPARTMENT OF DEFENSE LABORATORIES AND TEST ACTIVITIES
(FIGURES IN PARENTHESES ARE FOR THE R+D LABORATORIES)

PROGRAM DATA FOR FY 1969

	(MILLIONS $)	
TOTAL ANNUAL LABORATORY PROGRAM	4,208	(2,596)
TOTAL IN-HOUSE PROGRAM	2,090	(1,329)
TOTAL RDT+E PROGRAM	2,632	(1,790)
TOTAL IN-HOUSE RDT+E	1,410	(952)
TOTAL ANNUAL OPERATING COST	599	(329)

PERSONNEL DATA (END OF FY 1969)

PERSONNEL	AUTHORIZED STRENGTH	TOTAL PHD	TOTAL PROFESSIONALS
MILITARY	33,271 (9,032)	1,098 (915)	7,469 (3,769)
CIVILIAN	89,001(60,619)	2,501(2,401)	28,902(24,171)
TOTAL	122,272(69,651)	3,599(3,316)	36,471(27,940)

PHYSICAL FACILITIES (END OF FY 1969)

LAND (THOUSANDS OF ACRES)----------	7,363	(1,239)
SPACE (THOUSANDS OF SQ FT)---------	94,295	(47,122)
COST (MILLIONS OF DOLLARS)---------	6,267	(2,329)

FIGURE 1.3

The R&D laboratories can similarly be categorized into two parts - the twenty-three medical laboratories and the fifty-six non-medical laboratories. The mission and orientation of the medical laboratories are basically different from those of the non-medical. The two groups are also quite different in the magnitude of their staffing and funding. As can be seen from Figure 1.4, most of the medical laboratories have less than two hundred professionals and have an in-house R&D program of less than five million dollars, whereas the majority of the non-medical laboratories exceed these numbers.

Number of Professionals (Hundreds)								
		16-20	12-16	8-12	6-8	4-6	2-4	0-2
NON-MEDS	Army (24)	1		1	3	6	3	10
	Navy (20)		2	2	2	7	4	3
	Air Force (12)				2	1	6	3
MEDS	Army (11)						1	10
	Navy (8)							8
	Air Force (4)						2	2

In-House Research and Development Dollars (Millions)								
		30-100	25-30	20-25	15-20	10-15	5-10	0-5
NON-MEDS	Army (24)	3	1		4	3	3	10
	Navy (20)	4	1	3	1	4	4	3
	Air Force (12)			2		1	5	4
MEDS	Army (4)					1		10
	Navy (8)							8
	Air Force (4)						1	3

FIGURE 1.4
DoD Laboratories (Fiscal Year 1968)

The medical laboratories are also different from the non-medical laboratories in the composition and proportions of their data elements; while constituting almost one-third of the sample, they have but one-tenth of the professionals and only one-twenty-fifth of the R&D dollars. They tend to be staffed more by military than civilian professionals; they have 28% of the military professionals, and 63% of the military professionals with degrees at the doctorate level. (In the non-medical laboratories, civilians make up seven-eighths of the professional staff, whereas in the medical laboratories, the civilians are in the professional minority.) They also tend to rely more heavily on their own facilities and capabilities; their ratio of out-of-house R&D is only one-third that of the non-medical laboratories; and in magnitude they account for less than 1% of all laboratory dollars spent out-of-house.

The non-medical laboratories range in size from several with less than a hundred professionals to a few with more than a thousand. Among the smaller laboratories are four whose orientation is quite different from that of the other non-medical laboratories; these are the Army Behavioral Sciences Laboratory, the Naval Personnel Research Activity, the Naval Personnel Program Support Activity (now the Naval Personnel R&D Laboratory), and the Air Force Human Resources Laboratory. The smallest in number of personnel was the Air Force Academy's Frank J. Seiler Laboratory; because of its small size (a total staff of less than forty people) and other atypical characteristics, it was ultimately omitted from the study.

In the remainder of the report, these various categories of the remaining seventy-eight laboratories are referred to as follows: The term "medical" refers to the twenty-three laboratories with a primary medical mission, and the term "non-medical" refers to the remaining fifty-five laboratories. The non-medical are further subdivided into the four personnel research/behavioral sciences laboratories and the fifty-one physical sciences/engineering laboratories.

1.4 Chronological Development and Limitations of the Study

The study of relationships between the peer ratings and the quantitative properties of laboratories was undertaken shortly after the peer rankings were obtained, and was initially conducted by Dr. Steve Smith while assigned from the Missile Command Laboratories at Redstone Arsenal to the Office of Laboratory Management.[1]

[1] The Director of Defense Research and Engineering sponsored an intern program whereby personnel from the DoD laboratories are assigned a tour of duty in particular areas of DDR&E. Dr. Smith was the fourth such person to work for a six-month period in the Office of Laboratory Management, the author, from the Naval Weapons Laboratory, Dahlgren, Virginia, was the fifth.

The seventy-eight DoD laboratories were divided into two parts: the twenty-three medical laboratories and the fifty-five non-medical laboratories. Ratings for the two categories of laboratories were computed as described in Chapter 2, and those for the non-medical laboratories were then correlated in a variety of ways with the elements of the fiscal year 1968 data. Statistical analyses and computational support were provided by the staff of the Army's Harry Diamond Laboratory. Programming and computer support were also obtained from the Air Force's Information Systems Division in the Pentagon.

The present phase of the study, that which directly involves the author, began in April of 1970 shortly after his assignment to the Office of Laboratory Management. Dr. Smith had initiated the computation of correlations between the peer ratings and the fiscal year 1969 data. These were separated into the medical and non-medical components as was done previously with the 1968 data, but were further broken down according to the three military departments.

The separation of the DoD laboratories according to their respective military departments revealed distinctive differences in the correlations between the ratings and the quantitative data; the Navy laboratories exhibited some correlations that were half again as large as those for the DoD laboratories as a whole. A prime example was a correlation of .904 between peer rating and civilian professionals with master's degrees in eighteen non-medical Navy laboratories. An even higher correlation (.936) was observed between peer rating and the level of funds received from non-DoD sources in eleven Navy non-medical laboratories. Glass [3] offers an explanation for these higher correlations, noting that "most Navy Labs are strongly oriented toward engineering, and M.S. degrees for engineers are generally more indicative of the level of advanced training than doctorates. The ability of a laboratory to attract funds from sources other than its parent Military Department and other DoD components may be quite significant, since many other options are open to the non-DoD sponsor".

Based partially upon these results, and partially upon a desire to reduce the scope of the study, it was decided to exclude the medical laboratories from the present consideration. An influencing factor was that the results of the present work might be applicable to an evaluation of Project REFLEX [5] which was at that time just getting underway and which involved only physical sciences and engineering laboratories. Further, in addition to the significant differences in mission and size as noted in Section 1.3, it was obvious that the medical laboratories were not as well known as the non-medical laboratories. Although constituting almost one-third of the laboratories in number, they received less than one-eighth of the rankings. Two-thirds

DOD PHYSICAL SCIENCES AND ENGINEERING LABORATORIES

ARMY LABORATORIES

DA0001	LIMITED WAR LAB., ABERDEEN P.G., MD.
DA0004	ENGINEER TOPOGRAPHIC LAB., FT BELVOIR, VA.
DA0005	ENGINEER WATERWAYS EXPERIMENT STA., VICKSBURG, MISS.
DA0021	AERO. RES. LAB., AMES RES. CTR., MOFFETT FIELD, CAL.
DA0022	BALLISTIC RES. LABS., ABERDEEN PROVING GROUND, MD.
DA0023	COATING AND CHEMICAL LAB., ABERDEEN PROVING GROUND, MD.
DA0024	TERRESTRIAL SCIENCES CENTER, HANOVER, N.H.
DA0025	HARRY DIAMOND LABS., WASH., D.C.
DA0026	HUMAN ENG. LABS., ABERDEEN PROVING GROUND, MD.
DA0027	MATERIALS AND MECHANICS RESEARCH CENTER, WATERTOWN, MASS.
DA0028	NUCLEAR DEFENSE LAB., EDGEWOOD ARSENAL, MD.
DA0029	NATICK LABS., NATICK, MASS
DA0030	AVIATION MATERIEL LABS., FT. EUSTIS, VA.
DA0032	ELECTRONICS LABS., FT. MONMOUTH, N.J.
DA0033	MISSILE COMMAND LABS, REDSTONE ARSENAL, ALABAMA
DA0034	MOBILITY EQUIP. RES. AND DEVELOP. CENTER, FT BELVOIR, VA.
DA0035	FRANKFORD ARSENAL LABS., PHILA., PA.
DA0036	PICATINNY ARSENAL LABS., DOVER, N.J.
DA0037	BIOLOGICAL LABS., FT. DETRICK, MD.
DA0038	EDGEWOOD ARSENAL LABS., EDGEWOOD ARSENAL, MD
DA0039	ROCK ISLAND ARSENAL LABS., ROCK ISLAND, ILL
DA0040	WATERVLIET ARSENAL LABS., WATERVLIET, N.Y.
DA0041	TANK-AUTOMOTIVE COMMAND LABS., WARREN, MICH.

NAVY LABORATORIES

N00122	NAVAL UNWTR. WEAPONS RES. AND ENG. STA., NEWPORT, R.I.
N00156	NAVAL AIR ENGINEERING CENTER, PHILADELPHIA, PA.
N00167	NAVAL SHIP RESEARCH AND DEVELOPMENT CTR., WASH., D.C.
N00173	NAVAL RESEARCH LABORATORY, WASHINGTON, D.C.
N00178	NAVAL WEAPONS LABORATORY, DAHLGREN, VA.
N00953	NAVAL ELECTRONICS LABORATORY CENTER, SAN DIEGO, CALIF.
N60530	NAVAL WEAPONS CENTER, CHINA LAKE, CALIF.
N60921	NAVAL ORDNANCE LABORATORY, WHITE OAK, MD.
N61331	NAVAL MINE DEFENSE LAB, PANAMA CITY, FLORIDA
N61339	NAVAL TRAINING DEVICES CENTER, ORLANDO, FLA.
N61533	NSRDC(ANNAPOLIS) - (MARINE ENGINEERING LAB)
N62269	NAVAL AIR DEVELOPMENT CENTER, JOHNSVILLE, PA.
N62399	NAVAL CIVIL ENGINEERING LABORATORY, PORT HUENEME, CAL.
N62462	NAVAL APPLIED SCIENCE LABORATORY, BROOKLYN, N.Y.
N62479	NAVAL RADIOLOGICAL DEFENSE LAB., SAN FRANCISCO, CALIF.
N62738	NAVAL WEAPONS CENTER CORONA LAB, CORONA, CAL.
N63198	NAVAL UNDERSEA WARFARE CENTER, PASADENA, CALIF.
N70024	NAVY UNDERWATER SOUND LABORATORY, NEW LONDON, CONN.

AIR FORCE LABORATORIES

DF0105	MATERIALS LAB., WPAFB, OHIO
DF0121	AEROSPACE RESEARCH LABORATORIES, WPAFB, OHIO
DF0124	CAMBRIDGE RESEARCH LABS., L.G. HANSCOM FLD., MASS.
DF0301	ROME AIR DEVELOPMENT CENTER,GRIFFIS AFB, N.Y.
DF0303	AVIONICS LAB., WPAFB, OHIO
DF0506	WEAPONS LAB., KIRTLAND AFB, N.M.
DF0602	AEROPROPULSION LAB., WPAFB, OHIO
DF0604	FLIGHT DYNAMICS LAB., WPAFB, OHIO
DF0608	ROCKET PROPULSION LAB., EDWARDS AFB, CALIF.
DF0807	ARMAMENT LAB., EGLIN AFB, FLA.

FIGURE 1.5

of the participants mentioned no more than two medical laboratories, and half of these ranked none at all (see Section 2.3). In consequence of the smaller number of rankings, the ratings of the medical laboratories are not as significant as those of the non-medical laboratories.

For similar reasons, it was also decided to exclude the personnel research and the behavioral sciences laboratories. Like the medical laboratories, their mission is principally non-engineering oriented, their size characteristics are generally small compared to the physical sciences and engineering laboratories, and they received relatively fewer rankings.

The remaining fifty-one laboratories - the physical sciences and engineering laboratories - are shown in Figure 1.5. The remainder of this report is devoted to a discussion of the ratings and properties of these laboratories, and the relationships between them.

2. THE PEER RATING

2.1 The Peer Rankings

The peer rating survey was conducted as described in Section 2 of Chapter 1. Each participant was given a deck of cards containing the names of the laboratories, and was asked to separate the cards into two piles - those that he knew something about (enough to form a basis for rating), and those that he did not. He was then asked to separate those he knew into three groups - Above Average, Average, and Below Average - and to rank them within each group. Finally, the groups were combined in ranked order and the cards consecutively numbered from first to last.

Rankings were received from three hundred and twenty-five raters. The number of laboratories ranked by any one rater ranged from three to sixty. On the average, each rater ranked twenty-three laboratories; the median number was twenty-two. The distribution of rankings is shown in Figure 2.1. The vertical scale shows the number of participants who ranked the number of laboratories shown on the horizontal scale (0-4, 5-9, 10-14, ···, 55-60).

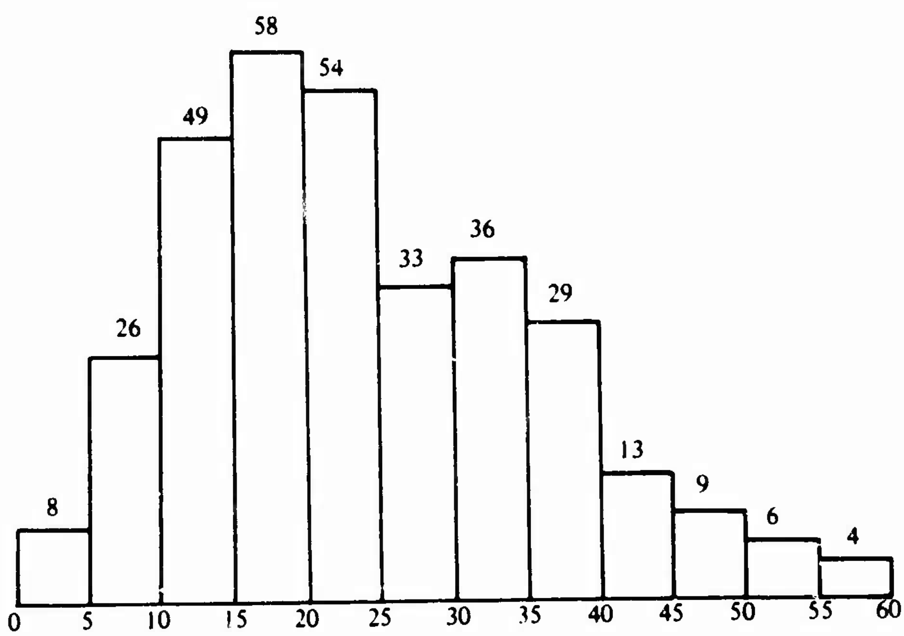


FIGURE 2.1
Distribution of Rankings

The distribution of rankings among the various rater groups is shown in Figure 2.2. Seventy-five percent of the raters were from within the Department of Defense, the majority of these being commanding officers and technical directors of the individual laboratories. Most of the remaining twenty-five percent were from private industry.

The distribution of rankings received by the laboratories is shown in Figure 2.3. The medical and personnel laboratories received an average of thirty-eight rankings each.[1] The rankings received by individual laboratories ranged from eleven to one hundred and four, with a median of thirty-two.

The average number of rankings of the physical sciences and engineering laboratories was one hundred and twenty-eight - more than three times as many as the medical and personnel laboratories. The range of rankings for the physical sciences and engineering laboratories varied from forty-five to two hundred and fifty, with a median of one hundred and twenty-four.

For the separate military departments, the average number of rankings per laboratory was 116 for the Army, 138 for the Navy, and 139 for the Air Force. Variations by rater groups and by military department are discussed further in Section 2.5, following a review of the method of computing the laboratory ratings.

2.2 Computation of Laboratory Rating

Since most of the raters ranked a different number of laboratories, a standardization procedure was required before the rankings of the various raters could be combined. In the work done by Dr. Smith, the standardization was accomplished by apportioning each ranker's opinions into ten equal zones (deciles), as was done by Apstein [1]. For example, if a participant rated twenty laboratories, the first two would be assigned a value of 10, the next two a value of 9, etc., with the last two being assigned a value of 1. The overall rating for each laboratory was then computed by averaging the number of votes received in each decile multiplied by the value of the decile. Relative standings based on the decile computations were communicated to each of the laboratories showing how they rated overall and within their own military department, and how their relative rank varied according to different categories of raters.

[1] The Frank J. Seiler Laboratory was also included in this group.

Type	Number of Rankers	Total Rankings	Average Number Ranked	Number Who Ranked From 1-9	10-19	20-29	30-39	40-49	50-60
1 OSD	44	1110	25	5	11	11	12	4	1
2 Service Headquarters	19	513	27	-	5	6	6	1	1
3 Service Commands	45	1133	25	1	19	11	7	3	4
4 DoD Laboratories	136	3225	24	11	42	42	30	11	-
5 Other Govt. Laboratories	6	133	22	1	2	2	-	-	1
6 Private Industry	66	1310	20	15	25	11	9	3	3
7 Academic and Not-for-Profit	9	181	20	1	3	4	1	-	-
ALL	325	7605	23	34	107	87	65	22	10

FIGURE 2.2
Rankings per Rater Type

Number of Laboratories Receiving 11-30, 31-50, ···, 231-250 Rankings

Mid-point of Class Interval	20	40	60	80	100	120	140	160	180	200	220	240
Medical and Personnel Laboratories	13	8	4	2	1	-	-	-	-	-	-	-
Physical Sciences and Engineering Laboratories	-	1	3	7	10	8	6	9	1	2	2	2

FIGURE 2.3
Rankings per Laboratory Type

In the present work, the standardization process is similar, except that instead of assigning the rankings to deciles, the rank for each laboratory is converted to a fraction (as described in the following section) and the rating is then the average of these fractions. This procedure was adopted to provide flexibility in experimenting with variations in the method of computing the ratings.

In either case, there is an underlying assumption that the rankings may be treated as if they were uniformly distributed from the highest to the lowest. This assumption uniformity is probably the most sensitive part of the rating procedure, since actually what was obtained from the survey is some unknown combination of rankings and ratings. This is because the participants were asked to first rate the laboratories by assigning them to the various groups (Above Average, Average, and Below Average) and then to rank them within the groups. However, since the various groups were re-combined by the ranker before turning his rankings in, it is not known -- except in ten instances enumerated in Figure 2.4 -- how the rankings were actually distributed among these groups. Nor is there any indication from the rankings about their spread, i.e., there is no scale to indicate how much better a participant considered one laboratory than another.

#1	#2	#3	#4	#5
+ Avg 3	+ Avg 3	+ Avg 2	+ Avg 21	A 18
– Avg 5	– Avg 2	– Avg 2	– Avg 19	B 23

#6	#7	#8	#9	#10
+ Avg 9	+ A 6	A 4	1 9	A 7
Avg 11	A 12	B 6	2 8	B_1 12
– Avg 13	– A 10	C 1	3 8	B_2 11
				C 8

FIGURE 2.4
Distribution of Rankings of Ten Raters

The notation accompanying the data presented in Figure 2.4 was taken from the raters' cards. Only one of the ten raters divided the middle group into two parts as instructed, and only four others even had a middle group. Raters 4, 5, 6, 7, 9, and 10 ranked more laboratories than the average rater, one wonders if these ratings are typical of the raters as a whole. (In the 1963 survey, Apstein observed that of an average of twenty-five votes per rater, six or seven were in the above average category, twelve were in the average category; and the remainder were rated below average.)

The validity of the assumption that the rankings are linearly distributed from high to low would seem to be most suspect when the number of rankings is small. For example, did a rater who ranked ten laboratories - and rated as ninth and tenth the two laboratories that the population as a whole considered the best and second best - consider his last two "below average", or were they "above average" and just at the end of his list?

Such speculation is fruitless in the present case. Empirical evidence is available from the people who have conducted such surveys and have had first-hand experience with the distribution of rankings. From their observations in conducting such surveys, Apstein, Anderson, and Glass have concluded that the distribution of peer rankings is statistically uniform, i.e., the individual deviations generally balance each other out.

2.3 Present Study; Modification of the Procedure

In the study conducted by Apstein, and in the previous phases of the present study conducted by Dr. Smith, the computation of laboratory ratings was performed by transforming the rankings to deciles. The present phase of the study modified the methodology in two ways: (1) the rankings were converted to fractions; and (2) not all of the rankings were used.

It was felt that converting the rankings to fractions instead of assigning them to deciles would provide a simpler mechanism for experimenting with different rating procedures, and would also help to preserve the rank-ordering inherent in the original data.[1] Various transformation models were considered; the one that was finally used was of the form

$$S = 10 \cdot (N - R) / (N - 1)$$

where N is the number of laboratories ranked by the particular rater, R is the ranking of a particular laboratory ($R = 1, 2, \cdots, N$), and S is the standardized score

---

[1] I had in mind to compute the ratings by weighting the rankings in various ways. For example, given thirty rankings, rather than apportion them three per decile, perhaps assign the first one the value 10, the next two the value 9, the next three the value 8, the next four the value 7, and the next five the value 6; then reverse the procedure to the low end of the scale.

of that laboratory for that rater. The overall rating for each laboratory is then simply the average of its standardized scores.[1] The ratings are generally lower than those computed by the weighted decile method, since in the present method the weight would tend to be near the mid-range of a decile (if deciles were used), while in the former case the weight was concentrated at the upper end of the decile.

The second change in procedure from that used previously by Apstein and Smith was the setting aside of some of the rankings. This came about in two ways: one was the introduction of a "threshold" value which required that a participant rank at least a minimum number of laboratories; the other was the utilization of relative rankings resulting from the exclusion of the medical laboratories.

The establishment of a threshold was done partly to assure that the ranker was familiar with a sufficient number of laboratories and partly to ensure that the rankings would be uniformly distributed. The figure ten was chosen as a threshold value; it might just as well have been nine or eleven; but a number such as five or four seemed too low in light of the procedure being used. The use of a threshold of ten excluded 34 raters. The decile equivalents of the rankings of the medical and personnel laboratories, using the standardized scores of the remaining 291 participants, are shown in Figure 2.5. (The top-rated medical laboratory received thirty-seven "first-place" votes, eighteen "second-place" votes, etc.) The rankings of the physical sciences and engineering laboratories are shown in Figure 2.6; these have been drawn to the same scale, but are based on a slightly reduced set of rankings, as described in the next paragraph. The laboratories are presented in the order according to which they were rated; for Figure 2.6, the order proceeds across the top row of both the left and righthand pages, then to the second row, etc. The patterns for the highest- and lowest-rated laboratories generally show well-defined modes, whereas some of those in the middle "grey" area are bi-modal or multi-modal, indicative of the wider range of variation of opinion concerning their technical competence. Some of the variation in modes is attributable to the artificial representation by deciles. It is obvious that the ratings of the medical and personnel laboratories are based on far fewer votes than are the ratings of the physical sciences and engineering laboratories.

---

[1] The model initially used was of the form $S = 10 \cdot (N - R + \frac{1}{2}) / N$. This gave fractional scores ranging from $1/2N$ to $1 - 1/2N$, whereas the model finally used gives fractional scores ranging from 0 to 1, which was considered preferable. However, in some instances, viz. Figures 2.12 and 2.15, it has been desirable to illustrate the distribution of the rankings in a bar-chart format with the scores grouped by deciles. In these circumstances the model described in this footnote has been used to generate the ratings because it results in a more natural distribution of scores to deciles. (The rank orderings generated by the two models are essentially the same. In the few instances where the two sets of rankings are not in item-by-item agreement, less than a 1% change in the rating of either one laboratory or the other would suffice to put them back into corresponding rank-order.)

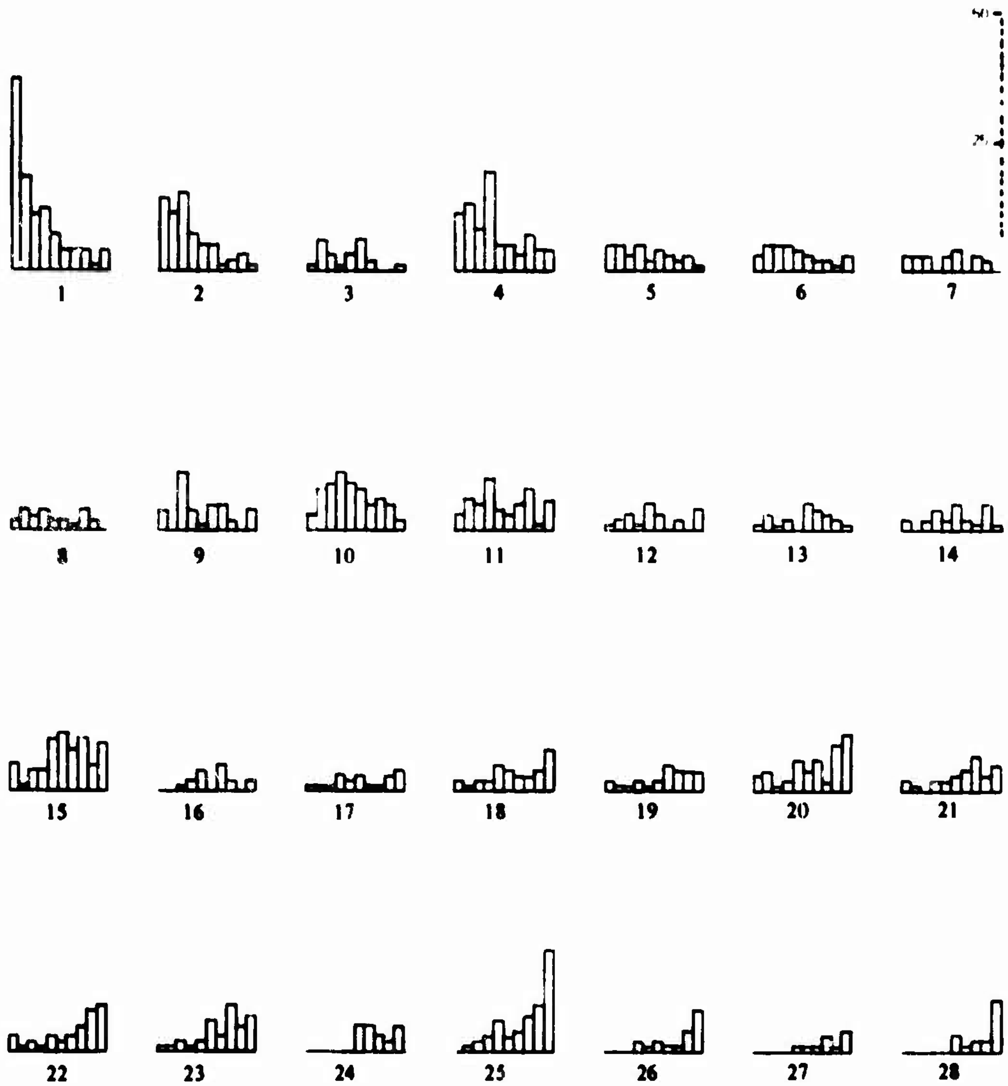


FIGURE 2.5
Distribution of Rankings of Medical and Personnel Laboratories by Deciles

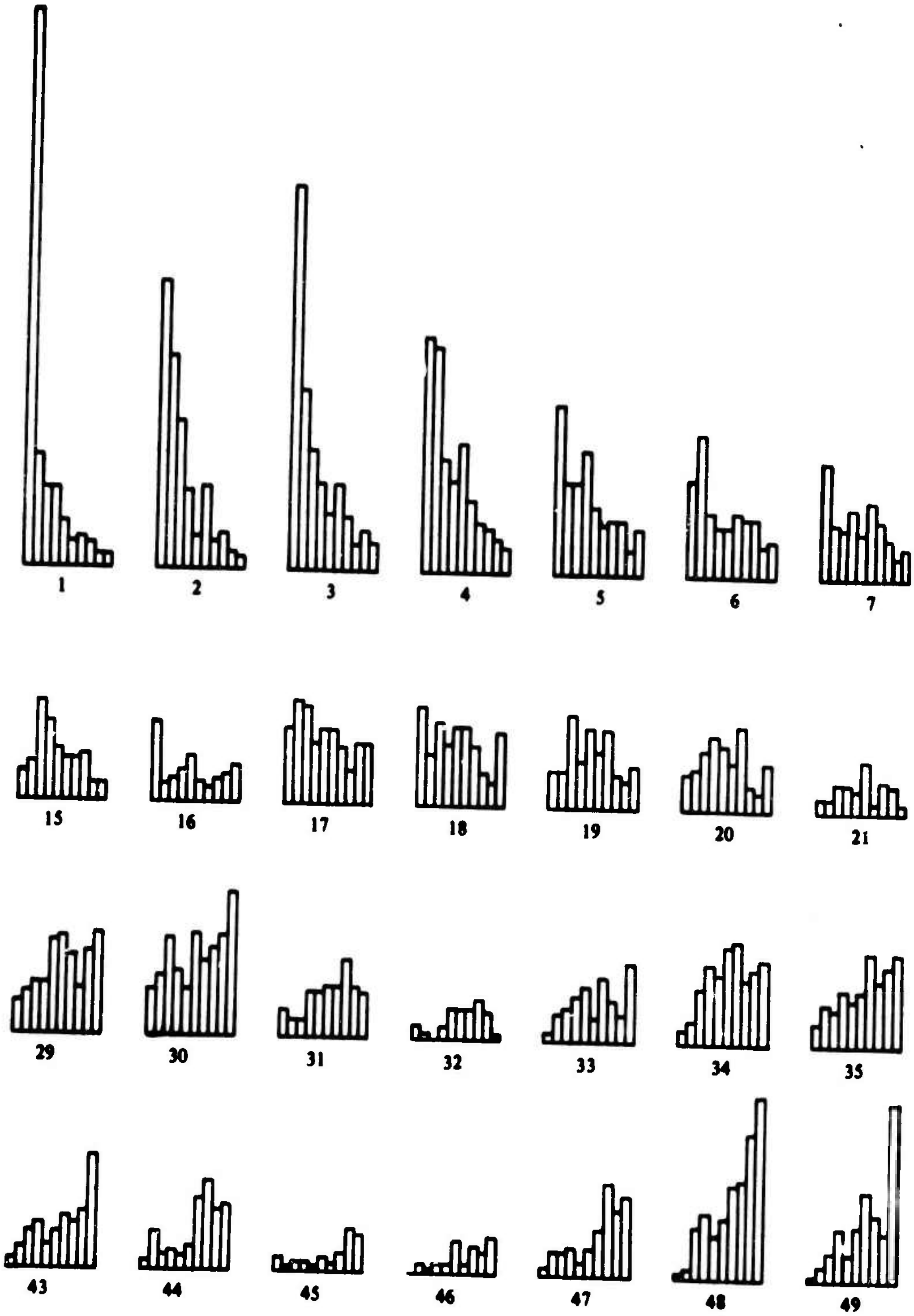
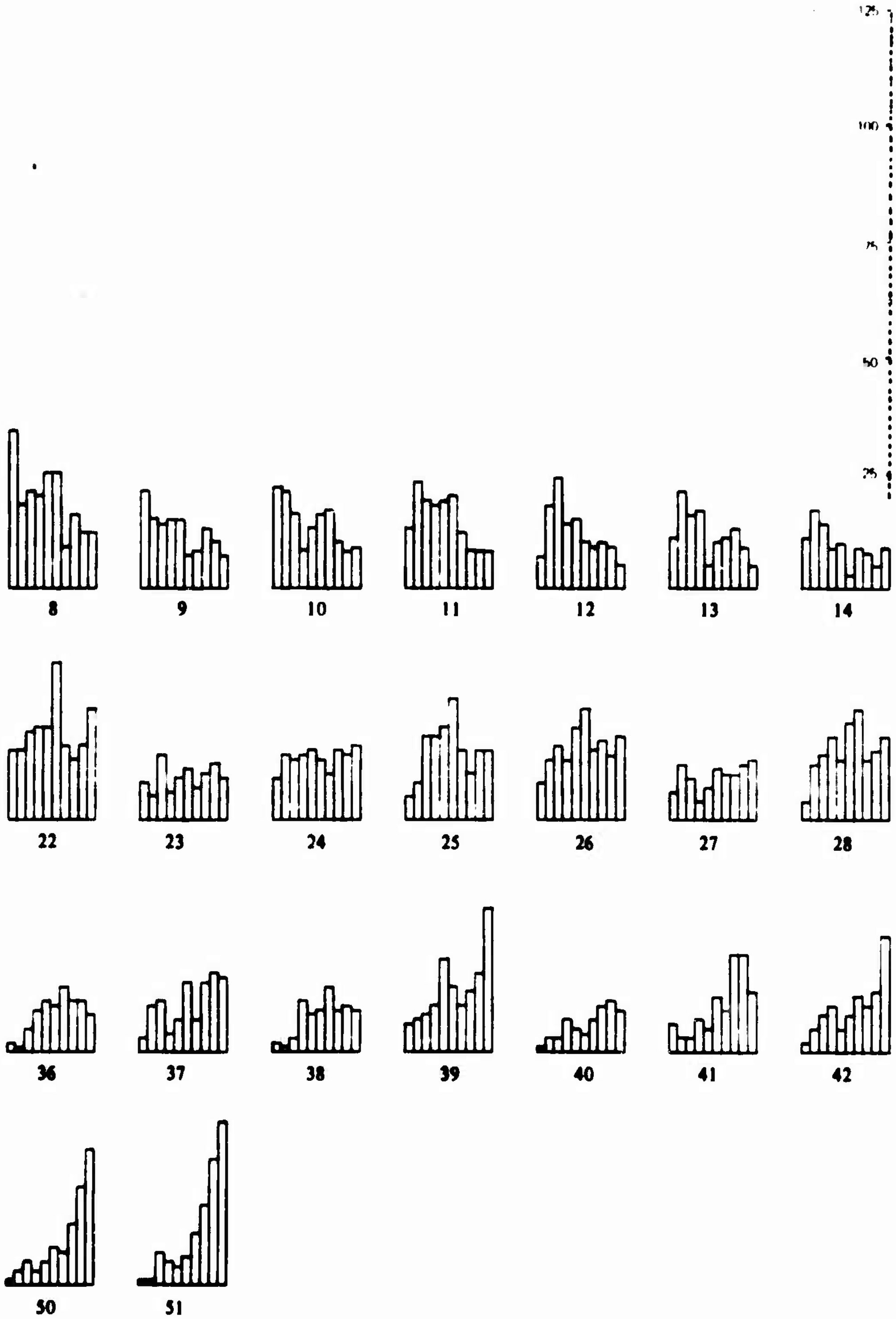


FIGURE 2.6
Distribution of Rankings of Physical Sciences and Engineering Laboratories by Deciles


125
100
75
50
25
8
9
10
11
12
13
14
22
23
24
25
26
27
28
36
37
38
39
40
41
42
50
51

Use of Relative Rankings

Following the decision to limit the present study to an examination of the non-medical laboratories, (see Section 1.4), it was decided to further restrict the sample to those participants who ranked ten or more non-medical laboratories, and in rating the non-medical laboratories, to include only the relative rankings of the non-medical laboratories. For example, if a particular rater had ranked thirteen laboratories -- ten non-medical and three medical -- only the ten non-medical rankings were used. If the medical laboratories had been ranked in positions 1-2-3, these were skipped over, and the non-medical laboratories were assigned ranks 1-10. This was possible to do because the major portion of the rankings were for the non-medical laboratories. The reverse procedure -- to rate the medical laboratories using only the rankings of the medical laboratories -- would have been much more suspect because of the paucity of their rankings.

These limitations resulted partially from an intuitive feeling that the study should be so conducted, and partially from observation. Since for the immediate study it was desired to obtain the relative rating of the physical sciences and the engineering laboratories, it was considered preferable to eliminate rankings that might be indicative of a medical bias or which might otherwise disturb the rating of the non-medical laboratories. Figure 2.7 shows some of the characteristics relevant to the distribution of ratings of medical and non-medical laboratories; notice that only 20% of the participants ranked five or more medical laboratories, and 40% none at all.

The requirement to rank at least ten non-medical laboratories further reduced the number of participants to 280. Seven of the eleven participants who were thus additionally removed had ranked more medical than non-medical laboratories; only three of the remaining 280 had this property. Of the four other participants removed, two ranked about an equal number of medical and non-medical laboratories, and the other two were already marginal with respect to the threshold, having ranked only 10 and 11 laboratories, respectively.

The rankings of the medical and non-medical laboratories in the remaining sample were fairly uniformly distributed by quartiles. Those who ranked five or more medical laboratories ranked about twice as many non-medical as medical laboratories in each of the four quartiles, as shown below.

	I	II	III	IV
Medical	8%	9%	9%	8%
Non-Medical	17%	17%	17%	17%

Number of Medical Laboratories Rated	By Those Who Rated < 10 NM	Cumulative Sum	By Those Who Rated ≥ 10 NM	Cumulative Sum	All Raters	Cumulative Sum
0	25	25	109	109	134	134
1	6	31	42	151	48	182
2	2	33	35	186	37	219
3	2	35	21	207	23	242
4	-	35	19	226	19	261
5	1	36	18	244	19	280
6	-	37	11	255	11	291
7	2	38	2	257	4	295
8	1	39	2	259	3	298
9	1	40	5	264	6	304
10	-	40	3	267	3	307
11	1	41	2	269	3	310
12	1	42	-	269	1	311
13	-	42	1	270	1	312
14	-	42	2	272	2	314
15	-	42	1	273	1	315
16	1	43	2	275	3	318
17	-	43	1	276	1	319
18	-	43	-	276	-	319
19	-	43	1	277	1	320
20	-	43	1	278	1	321
21	1	44	-	278	1	322
22	1	45	-	278	1	323
23	-	45	2	280	2	325
		45		280		325

FIGURE 2.7
Distribution of Raters According to Number of Medical Laboratories Rated

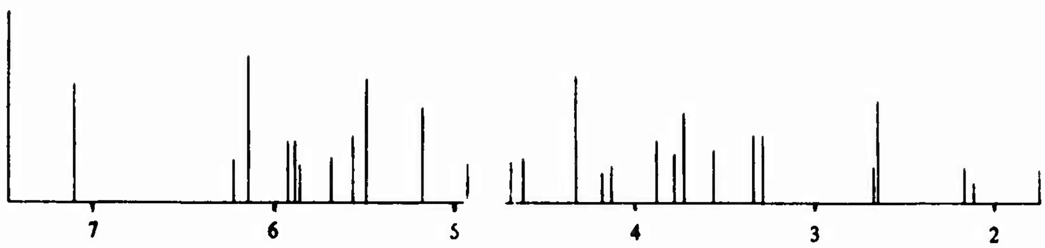
The investigation of the effects of the rankings of the medical laboratories upon the ratings of the non-medical laboratories raised the question of whether or not to similarly exclude the personnel laboratories. The basic question was whether or not the personnel laboratories were coupled with the medical laboratories, or whether their rankings were independent of the medical laboratories. The evidence indicates that there is some coupling between the medical and the personnel laboratories. More than 85% of those who ranked a personnel laboratory also ranked at least one medical laboratory, compared to less than 50% of those who did not. More than 45% of those who ranked a personnel laboratory ranked at least five medical laboratories, compared to less than 10% of those who did not. Despite the apparent coupling between the rankings of medical laboratories and the rankings of personnel laboratories, the personnel laboratories were rated with the other non-medical laboratories. It is doubtful that their removal would have made any significant difference in the ratings overall.

The distribution of the laboratory ratings is shown in Figure 2.8. The ratings of the medical and personnel laboratories are based on the 291 raters who ranked ten or more laboratories; the ratings of the physical sciences and engineering laboratories were computed from the rankings of the 280 participants who ranked ten or more non-medical laboratories. The horizontal axes show the distribution of ratings of the various laboratories; the vertical axes are proportional to the number of rankings upon which the rating was based. The figures suggest that the more times a laboratory is ranked, the higher will be its rating; the correlation between the ratings of the physical sciences and engineering laboratories and the number of rankings is .6.

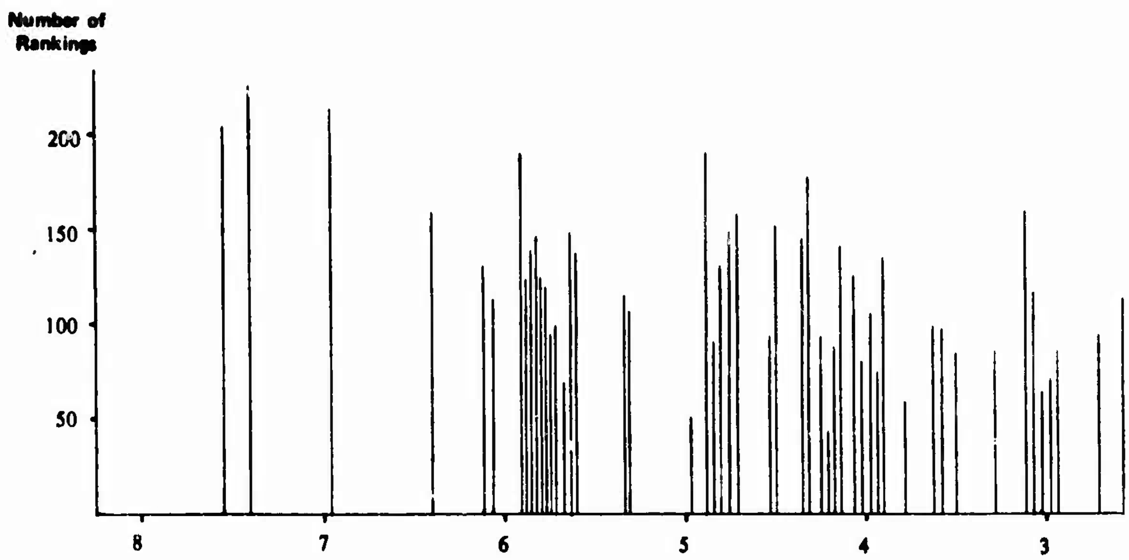
The ratings obtained in this way, using the rankings of the participants who ranked at least ten non-medical laboratories, were adopted as the "standard" ratings and are the principal ones used in the later chapters in looking for associations between the peer ratings and the laboratory properties.

2.4 Realiability of the Ratings

Questions about the ratings ususally fall into two categories: what are they really measuring, and what is their statistical significance. This section addresses only the latter question. In this sense the term "reliability" is concerned with how the ratings vary with the method of computing; how they depend upon the size of the threshold; how they are affected by biases among the different rater groups; and to what degree they are significantly different from one another.



Ratings of
Medical and Personnel Laboratories



Ratings of
Physical Sciences and Engineering Laboratories

FIGURE 2.8
Distribution of Ratings of DoD Laboratories (Horizontal Scale) vs Number of Rankings per Laboratory (Vertical Scale)

Consistency and Stability of the Ratings

"The score of a laboratory ... represents an average of the opinions of all participating raters. This score in itself unfortunately tells little about the consistency of opinion among the raters. If ten individual raters were in complete disagreement about a given laboratory, there would be one assignment to each of the ten deciles. This would yield a score of 5.5. If there were essential agreement among then with five assignments to the sixth decile and five assignments to the fifth decile, the score would still be 5.5. One method of gaining an impression as to the consistency among the various raters is by directly observing the number of raters assigning a laboratory to each of the ten deciles. However, it would be convenient if one number could be used to compare the consistency of opinion about a laboratory. The standard deviation (σ) is often used for this purpose. A low standard deviation indicates high consistency.

"The average score and the standard deviation indicate nothing about the stability of the computation. That is, were there enough raters to give a reasonable estimate of this average opinion? Little can be said about this question without distributional assumptions. Assuming normality, a 95% confidence interval ($\bar{x} \pm \gamma$) for the mean can be calculated. This means that a "true" mean score further away than the calculated 95% confidence interval from the previously calculated mean score would be unlikely (i.e., have a probability of less than 0.05). Thus if the calculated value of the confidence interval is small, in our case less than 1.0, then the mean score is a stable value."[1]

The standard deviations and confidence intervals of the ratings of the physical sciences and engineering laboratories, computed form the rankings of those who ranked ten or more non-medical laboratories, are shown below.

	σmin	σmax	$\bar{\sigma}$	γmin	γmax	$\bar{\gamma}$
Army	2.22	3.21	2.7	0.33	0.77	0.5
Navy	2.28	2.78	2.5	0.30	0.56	0.5
Air Force	2.55	2.91	2.8	0.43	0.52	0.5

The 95% confidence intervals for the standard ratings are depicted in Figure 2.9. The average interval is about one decile in width (.5 on each side of the rating). The overlapping zone formed by the lower bound of one confidence interval and the upper bound of another spans an average range of about thirteen laboratories. The smaller intervals within the confidence intervals reflect variations in the ratings according to the number of raters, as discussed below.

[1]From Dr. Smith's unpublished notes.

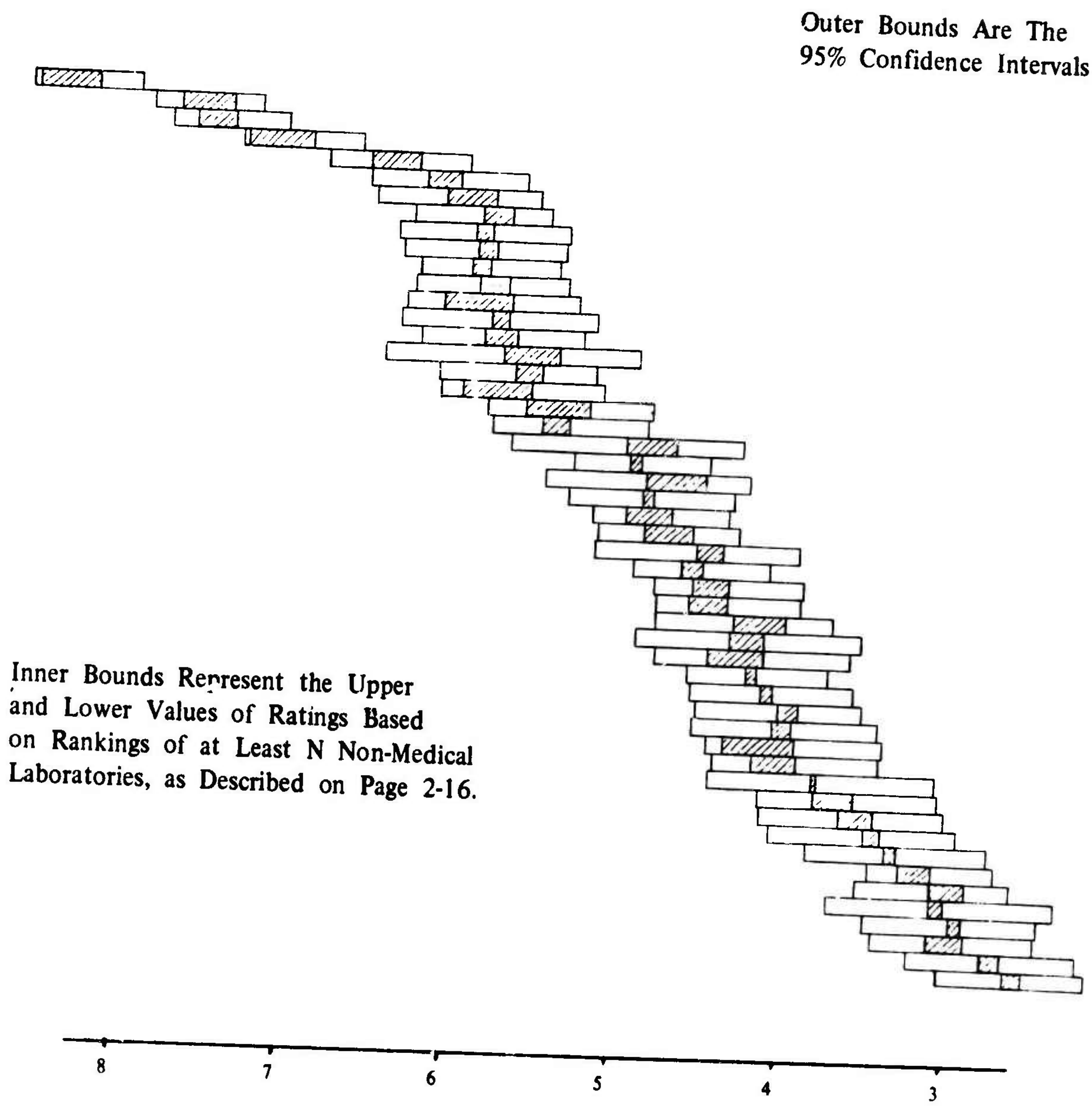


FIGURE 2.9
95% Confidence Intervals for Peer Ratings

Variation in Peer Rating With Number of Participants

In order to determine the sensitivity of the ratings to the limitation that the participants rank ten or more non-medical laboratories, the threshold value was also set equal to five, fifteen, and twenty. The effect for five was between the results for zero and ten; the effects of the other thresholds including zero, are shown in Figures 2.9 and 2.10. The left and right boundaries of the inner intervals depicted in Figure 2.9 correspond to the highest and lowest values of the ratings obtained from the five thresholds. The largest variation is about half the size of the 95% confidence interval (which itself was determined from the ratings obtained with a threshold of ten). In only one case does the span of the variation exceed 10% of the value of the laboratory's rating.

Figure 2.10 shows how the rank-ordering of the laboratories changes according to the various methods used.[1] The rank-orderings in each column are given relative to the standard column (Column 4). For example, the laboratory ranked twelfth in the standard ratings was ranked tenth for a threshold of zero and fourteenth for a threshold of fifteen. Column 1 shows the rankings obtained by partitioning the votes of all raters into deciles. The remaining columns use the scoring equation described in section 2.2. Column 2 gives the rankings using all votes; Column 3 is computed from the rankings who rated ten or more laboratories; and Columns 4, 5, and 6 are computed from the rankings of those who rated at least 10, 15, or 20 non-medical laboratories, respectively. The higher-ranked laboratories suffered a proportionally higher loss of raters as the threshold increased. The top seventeen laboratories averaged 34% less rankings for a threshold of twenty than for a threshold of zero; the middle seventeen decreased by 27%; and the low seventeen decreased by 22%.

The ratings shown in all six columns can be divided into four non-overlapping zones. The first five laboratories are always the first five; they also maintain their relative order except for a threshold of fifteen. The next fifteen laboratories - those rated sixth through twentieth - also form a group. A third group is composed of the nineteen laboratories numbered twenty-one through thirty-nine; the last group is made up of the laboratories rated forty through fifty-one.

The rankings of the laboratories in the first and fourth groups are relatively stable across the various thresholds, and except for one instance with a threshold of twenty, can be put in corresponding rank order with less than a one percent change in the ratings. The

[1] The product-moment correlation between the ratings and the rank-ordering of the ratings is .963.

Deciles	Conversion to Fractions				
All Raters	All Raters	Threshold 10	Threshold 10 NM	Threshold 15 NM	Threshold 20 NM
1	1	1	1	1	1
2	2	2	2	3	2
3	3	3	3	2	3
4	4	4	4	4	4
5	5	5	5	5	5
6	6	6	6	7	6
7	11	7	7	6	13
11	9	8	8	10	7
$\boxed{12}$	8	10	9	11	18
8	$\boxed{12}$	11	10	9	11
10	7	9	11	13	9
9	10	$\boxed{12}$	$\boxed{12}$	14	$\boxed{12}$
15	15	13	13	16	15
13	14	15	14	$\boxed{12}$	14
14	13	17	15	8	8
17	16	14	16	17	10
16	17	16	17	15	19
18	18	18	18	18	17
20	20	20	19	19	20
19	19	19	20	20	16
21	22	21	21	22	25
22	21	22	22	24	22
23	26	23	23	25	24
26	23	26	24	21	23
24	24	24	25	28	21
25	25	25	26	26	26
27	27	27	27	27	28
28	28	28	28	23	30
29	29	29	29	29	29
30	30	30	30	30	33
31	31	31	31	32	27
32	34	32	32	33	32
34	32	34	33	34	38
39	33	35	34	31	34
35	39	33	35	35	39
33	35	36	36	38	35
36	36	39	37	39	37
38	37	37	38	36	31
37	38	38	39	37	36
40	41	40	40	40	40
41	40	41	41	42	41
42	42	42	42	41	42
43	43	43	43	43	43
44	44	44	44	44	44
45	47	45	45	45	45
47	45	46	46	47	49
46	46	47	47	46	47
48	48	49	48	48	48
49	49	48	49	49	46
50	50	50	50	50	50
51	51	51	51	51	51

FIGURE 2.10
Variation in Rank-Order

most variation is in the second and third groups, principally for a threshold of twenty. The ninth-ranked laboratory becomes the eighteenth; the fifteenth advances to ninth position; the thirty-third becomes the thirty-eighth; and thirty-sixth moves up to thirty-first. The largest variations occur where the ratings are most closely bunched.[1] There is less than a 5% difference between the 8th and 18th ranked laboratories, and there is less than an 11% difference between the 29th and the 39th ranked laboratories.

2.5 Variation by Rater Groups

As noted in the introduction, the raters were classified into seven groups. Two of the groups - other federal laboratories and universities and not-for-profits - accounted for less than five percent of the total raters; these will not be considered here. Also for this discussion, the rankings of the Headquarters Staffs have been combined with those of the Service Commands, principally because of the relatively small number of participants in the Headquarters group. The other groups are from the staff of the Director of Defense Research and Engineering (DDR&E), the DoD Laboratories, and Private Industry.[2]

One way of looking at the differences (and similarities) between these four principal groups is to observe which laboratories were rated high and which ones low by the various groups. The rankings of the first twenty and the last twelve laboratories, using the standard ratings obtained in Section 2.2, are shown in Figure 2.11.[3] The numbers in Columns 2-5 refer to the laboratories as ranked in Column 1. For example, the laboratory rated twelfth overall was rated seventh by DDR&E, eleventh by Staff and Commands, fourteenth by Laboratories, and seventeenth by Industry. The numbers in the lower table show the number of times that the ratings were based on fewer than fifteen rankings. Three of the rankings shown in Figure 2.11 reflect ratings based on fewer than ten votes: the rating of the laboratory rated eighteenth by DDR&E is based on eight votes, and those of the laboratories ranked fortieth and forty-eighth by Industry are based on nine and three votes, respectively [these are the laboratories marked with asterisks].

---

[1]Glass shows a similar comparison for the laboratories of a single department in reference (2).

[2]The coefficient of rank-order correlation between the ratings of the Headquarters Staffs and the Service Commands was .81. For the ratings of the two groups combined, the corresponding correlation with the DDR&E ratings was .77, and .72 with the Industry ratings. The coefficient of rank-order correlation between the DDR&E ratings and the Industry ratings was .64.

[3]The change in rating over a span of any ten laboratories in the second group is less than 12%, whereas the average change in rating over equivalent spans for all fifty-one laboratories is 21%.

	All Raters	DDR&E	Staffs and Commands	Laboratories	Private Industry
Rankings of First Twenty Laboratories	1	1	3	1	2
	2	3	1	2	1
	3	4	4	3	9
	4	8	2	4	3
	5	2	18	6	17
	6	15	16	5	13
	7	12	5	15	14
	8	14	9	8	7
	9	5	10	7	4
	10	18	6	11	10
	11	7	12	21	19
	12	13	7	13	5
	13	11	15	10	23
	14	6	14	12	11
	15	25	20	19	16
	16	19	8	17	20
	17	10	28	14	12
	18	21*	11	9	39
	19	22	17	25	18
	20	20	22	16	8
Rankings of Last Twelve Laboratories	40	49	44	40	36*
	41	40	34	42	18
	42	44	38	44	46
	43	37	37	43	45
	44	50	48	45	51
	45	39	50	39	43
	46	43	41	48	42
	47	48	45	47	48
	48	42	46	46	32*
	49	47	51	49	50
	50	45	47	50	44
	51	41	49	51	33
Top Twenty	<15	4	1	0	3
Bottom Twelve	<15	8	4	0	7

FIGURE 2.11
Ratings of DoD Laboratories

Ten of the first twenty laboratories were mentioned by all four groups; the other ten of the top twenty were each mentioned by three groups. The first four laboratories overall were rated in the top ten of each group; the fifth was rated in the top ten of three groups, but did not make even the top twenty of the Industry groups. Each of the first eighteen laboratories in the overall rankings was mentioned at least once in the top ten of the various groups, indicating again the bunching of the ratings and the diversity of opinion among the raters.

There was slightly more unanimity of opinion about the ten lower-ranked laboratories; five of the last ten were ranked in the lower twelve laboratories of each group, and the other five were each mentioned by three groups. Each of the laboratories ranked in the last twenty of the overall ratings was mentioned at least once in the lower twelve ratings of each group, except for the laboratory rated in thirty-fifth position.

There seems to be more divergence between Industry and the DoD groups than there is among the three DoD groups themselves. Some of this may be due to the number of ratings, but a substantial part is probably attributable to a different point of view. For example, the laboratory ranked thirty-ninth overall was ranked eighteenth by Industry – based on thirty votes – and the laboratory ranked eighteenth overall was ranked forty-first by Industry – based on thirteen votes.

Figure 2.12 illustrates how the rankings of the various groups were apportioned to deciles for each of the military departments. This includes the rankings of all raters with a threshold of zero, i.e., no rankings have been discarded, and the rankings of the medical and personnel laboratories are also included. (Later on, in Figure 2.15, will be shown the rankings used in computing the ratings of the physical sciences and engineering laboratories.) The three rows in Figure 2.12 represent the rankings received by Army, Navy, and Air Force laboratories, respectively. The dotted lines show the average number of rankings that would have been received if the raters in the corresponding group had voted for a proportionately equal number of laboratories in each of the three services. The five columns represent the various rater groups, as follows:

Column 1: rankings by DDR&E
Column 2: rankings by Headquarters Staffs and Service Commands
Column 3: rankings by Laboratories
Column 4: rankings by Private Industry
Column 5: rankings by other groups

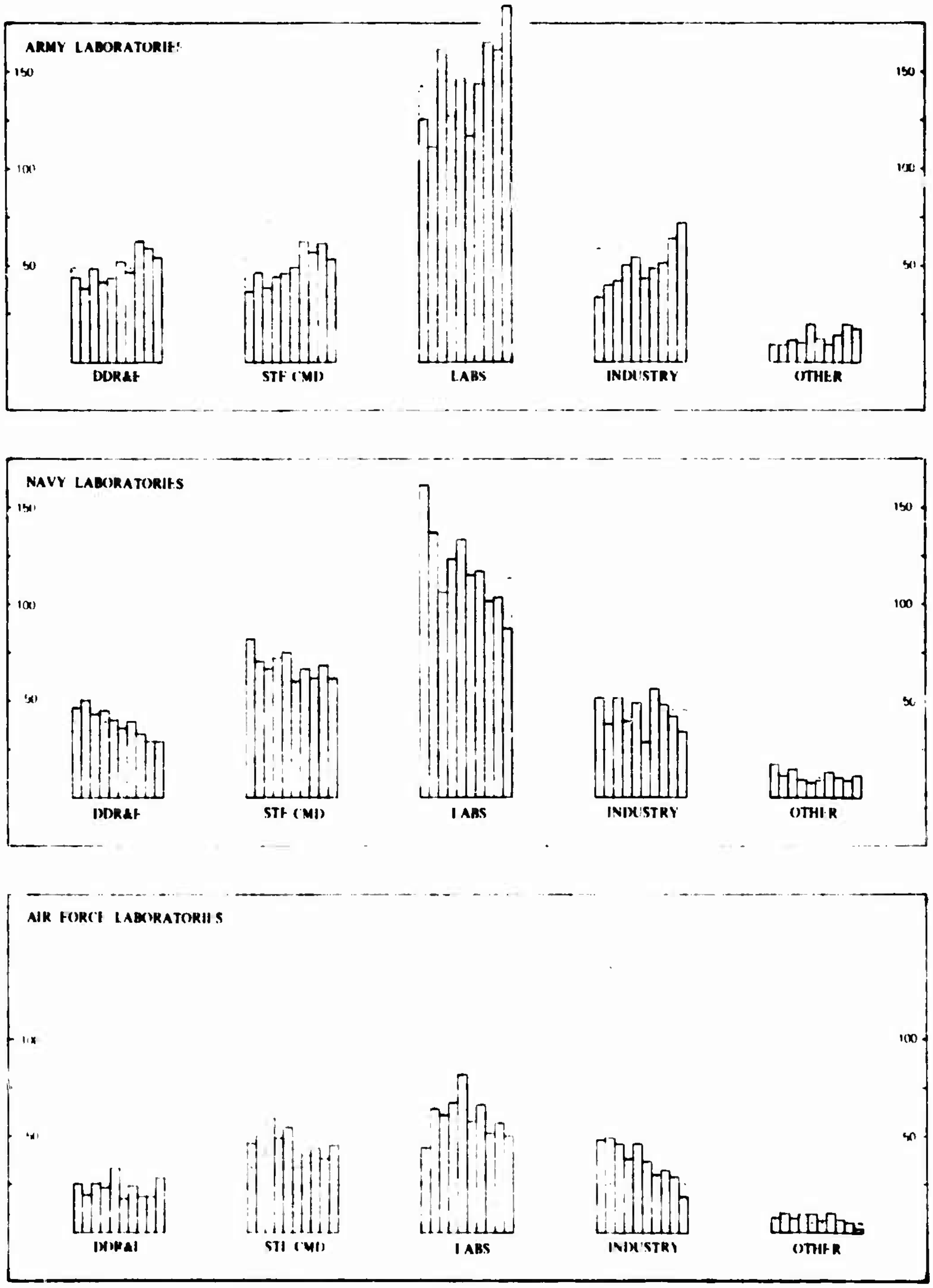


FIGURE 2.12
Distribution of Rankings by Service, According to Rater Groups

Variations by Military Departments

As part of a regression analysis conducted to determine which elements of the data base were most significantly correlated with the ratings (see Section 4.5), blocking variables were entered along with the elements to determine if there were signficant variations among the rater groups. When the rankings were blocked according to the standard rater groups, the differences between groups were found to be marginally significant; but when the rankings were blocked according to service affiliation, the ratings of the Army and the Air Force laboratories were found to be significantly dependent upon the composition of the rater groups.

Most of the DoD raters (other than DDR&E) can be identified according to parent military department. This is because the raters from Headquarters Staffs and Service Commands were asked to indicate their service affiliation; and although the laboratories were not so requested, many of them did (see Figure 2.14). The distribution of the rankings of the physical sciences and engineering laboratories by the one hundred and thirty-six raters with a known service affiliation is shown in Figure 2.13. The last table in the figure is derived from the one just above it (Average Number of Rankings...) divided by the number of physical sciences and engineering laboratories in each of the military departments (23, 18, and 10 for the Army, Navy, and Air Force); it is indicative of the percent of own and other service laboratories "known" to the raters in the three military departments. The "average" Army rater ranked three-quarters of the twenty-three Army laboratories versus one-third of those of each of the other services; the average Navy rater ranked more than four-fifths of the eighteen Navy laboratories, but less than one-fifth of those of each of the other two departments, and the typical Air Force rater ranked all ten Air Force laboratories and three-tenths each of the Army and Navy laboratories.

The distribution of rankings of the individual service raters follows a similar pattern. For example, of the twenty-nine raters identified with the Army laboratories, in all cases except one the number of rankings for Army laboratories exceeded the sum of rankings for the other two services, and for the thirty-four raters from Navy laboratories, the sum of rankings for Army and Air Force laboratories exceeded those for the Navy only twice. The rankings of Air Force raters were characterized by the density of votes cast for Air Force laboratories, twelve of the raters mentioned all ten Air Force Laboratories, and the thirteenth mentioned nine; only two of the raters in the laboratories of the other two services ranked as many as nine Air Force laboratories. In several cases there are strong indications of a preference to rate the laboratories in one's own department higher than those of the other departments; and in some instances there is more than a

Rankings By Service

Number of Raters	Service Affiliation	Army	Navy	Air Force	Total Rankings
38	Army	644	229	129	1002
58	Navy	250	892	101	1243
40	Air Force	259	213	392	864

Percent Distribution By Service Affiliation

	Army	Navy	Air Force
Army	64	23	13
Navy	20	72	8
Air Force	30	25	45

Average Number of Rankings By Service Affiliation

	Army	Navy	Air Force	Total Rankings
Army	17.0	6.0	3.4	26.4
Navy	4.3	15.4	1.7	21.4
Air Force	6.5	5.3	9.8	21.6

Percent of Department Laboratories Rated

	Army	Navy	Air Force
Army	74	33	34
Navy	19	85	17
Air Force	28	30	98

FIGURE 2.13
Rankings of DoD Laboratories
By Service Affiliation

suspicion that laboratory raters have placed proprietary pride above objectivity. How this affects the rating is uncertain, since the identities of the laboratories are unknown.[1]

The rankings of the 280 raters on which the standard ratings were based are shown in Figure 2.15. The outer figures in each row represent the rankings by DDR&E and Industry, as in Figure 2.12; the three inner figures are the rankings by Army, Navy, and Air Force raters, respectively. These include the known raters from the headquarters staff, the service commands, and the laboratories of each of the military departments, and in addition include twenty-six raters whose service affiliation was inferred from the pattern and distribution of their rankings. The composition of the raters is shown in Figure 2.14.

The diagrams show the extent to which the three services tend to rank only the laboratories within their own military department. As in Figure 2.12, the dotted lines indicate the proportion of rankings that would have been received if the raters within each group had mentioned a proportionately equal number of laboratories in each service. It is apparent from these figures that the overall DoD rank of a laboratory is highly dependent upon the composition of the sample. It indicates further, that unless the sample of raters can be arranged to uniformly represent the individual services, the services should be considered separately from one another in order to obtain a more realistic estimate of their relative technical competence.

Group	Army	Navy	Air Force	Total
DDR&E	--	--	--	37
Headquarters Staffs	1	7	9	17
Service Commands	8	17	18	43
Laboratories (Known)	29	34	13	76
Laboratories (Inferred)	18	7	4	29
Laboratories (Unknown)	--	--	--	14
Other Govt., Universities, etc.	--	--	--	13
Private Industry	--	--	--	51
	56	65	44	280

FIGURE 2.14
Distribution of Raters by Service Affiliation

---

[1] In his 1963 survey, Apstein found that recomputing the rating, leaving out the rankings of the judges attached to a particular laboratory, made less than five percent difference in the worst case, and in some instances actually improved the score for the laboratory. He concluded that "on the whole, the judges were harsher in judgement of their own installations than they were of others."

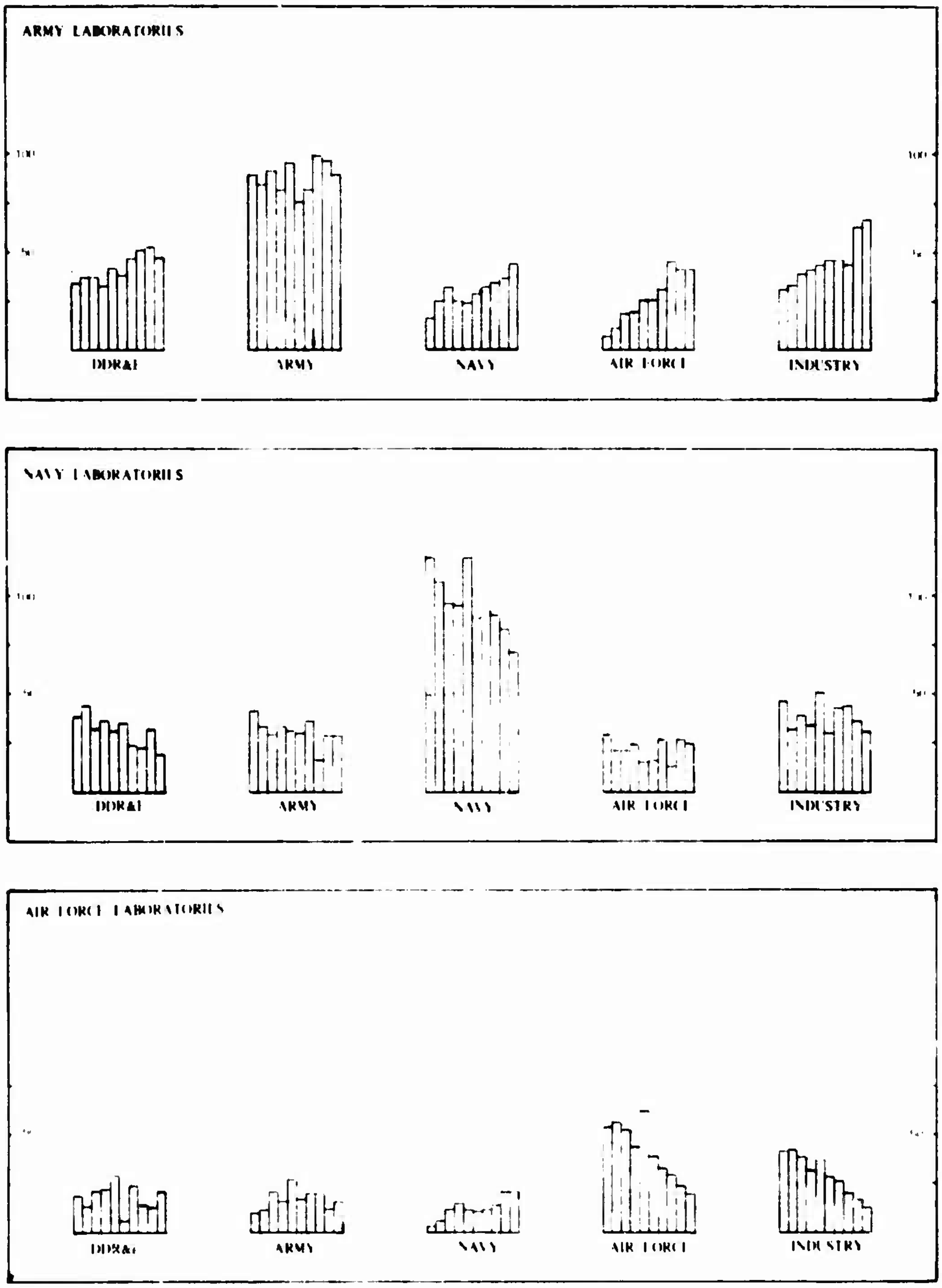


FIGURE 2.15
Distribution of Rankings by Service,
According to Rater Groups
(≥10 Non-Medical Laboratories)

Rankings Within Military Departments

In separating the ratings of the laboratories according to military department, it would be conceivable to adopt a procedure of relative ranking as was done in skipping over the medical laboratories in the computation of the standard ratings, i.e., in computing the ratings of the Army laboratories, to use only the rankings of the Army laboratories and to rank them 1,2,3,··· etc; however, this has not been done; the ratings have been computed following the standard procedure. The first three groups - program managers and technical specialists in the office of the Secretary of Defense and in the service headquarters and commands have been combined in order to provide a broader base. The rank-order of their combined ratings is shown under the column labeled "DSC" (for DDR&E, Staffs, and Commands). These groups differ from the other rater groups in that they control or influence the functions of the laboratories, as well as being users of the laboratories' services and products.

The overall rankings of the laboratories within each service are shown in Figure 2.16.[1] The first column is the ranking based on the standard ratings, the rankings in the other columns are identified relative to those in Column 1. For example, the Army laboratory ranked fifth overall is rated second by the DoD non-laboratory judges (Column 2), sixth by the laboratory raters (Column 3), and fourth by the raters from other Army laboratories (Column 4). There is some overlap in the rater groups, since the service raters include those from the headquarters staffs, the service commands, and the laboratories. The rankings by the industry raters have not been shown because in many cases there was an insufficient number (less than fifteen) of votes, although this was not true for the Air Force laboratories.

Statistical tests have not been applied to determine the significance of differences in the rankings shown in Figure 2-16, but a conservative rule of thumb would be to use a span of nine for the Army laboratories, eight for the Navy, and seven for the Air Force, i.e, to assume that the range of variation of sixth-ranked Army laboratory is distinct from range of variation of the sixteenth-ranked Army laboratory. The line graphs at the foot of each table of rankings indicate the distribution of the standard ratings for that department. The highest-rated laboratory is on the left; the lowest-rated one on the right. The numbers shown under the heading "Number of Raters" correspond to the maximum and minimum number of rankings upon which the ratings were based.

---

[1]The correlation between the ratings and their rank-order is .964 for the Army, .963 for the Navy, and .922 for the Air Force.

Ranking of Army Laboratories

All	1-2-3	Labs.	Army
1	1	1	1
2	$\boxed{5}$	2	2
3	2	6	3
4	3	3	$\boxed{5}$
$\boxed{5}$	4	4	9
6	7	$\boxed{5}$	4
7	10	9	6
8	9	11	10
9	6	13	8
10	8	8	13
11	13	7	7
12	11	12	11
13	12	14	14
14	14	15	12
15	15	10	17
16	16	16	16
17	17	17	15
18	18	18	18
19	22	19	19
20	21	21	21
21	19	20	20
22	20	22	22
23	23	23	23

Ranking of Navy Laboratories

All	1-2-3	Labs.	Navy
1	1	1	1
2	2	2	2
3	3	3	3
4	6	4	4
5	7	7	6
6	4	5	5
7	5	6	7
8	8	8	9
9	10	14	8
10	9	13	10
11	11	9	11
12	12	12	12
13	16	11	16
14	14	10	14
15	17	15	13
16	15	16	15
17	13	17	17
18	18	18	18

Ranking of Air Force Laboratories

All	1-2-3	Labs.	Air Force
1	1	1	4
2	4	2	1
3	2	5	2
4	3	4	3
5	8	7	6
6	5	6	8
7	6	3	5
8	7	8	9
9	9	9	7
10	10	10	10

Number of Raters

	Army All	Army 1-2-3	Army Labs.	Army	Navy All	Navy 1-2-3	Navy Labs.	Navy	Air Force All	Air Force 1-2-3	Air Force Labs.	Air Force
Max	206	68	95	54	235	78	102	62	160	57	61	44
Min	44	16	23	20	76	35	27	42	107	43	32	40

FIGURE 2.16
Variations in Ranking of DoD Laboratories According to Rater Subgroups

There was a considerable amount of bias present in the ratings by the service groups. For example, the Army raters ranked six Army laboratories in the first nine DoD laboratories; the Navy raters ranked seven Navy laboratories in the first eight; and the Air Force raters ranked seven Air Force laboratories in the first ten. The Army ratings of Army laboratories exceeded the standard ratings of the Army laboratories except for two of the last three cases. The Navy rated its first twelve laboratories higher than the corresponding ratings by the total sample; four of the last six were rated lower. The Air Force ratings were higher in all but two cases.

Overall, while there is considerable variation in the numberical ratings given by the various groups, the rankings of the top and bottom laboratories of the various services show a high degree of consistency. The top five Army laboratories are ranked among the first six Army laboratories in all four sets of rankings; the bottom five laboratories are always among the last five. The top three Navy laboratories are consistently ranked in that order; the bottom four appear in the last six of each of the arrangements shown. Three of the top four Air Force laboratories are always among the first four; the bottom three are in each case among the last four. The Air Force laboratory ranked in third place overall is ranked in seventh place by the laboratory raters; however, its rating is within ten percent of that of the one ranked third.

2.6 Comparison With Previous Studies

Twenty-nine of the fifty-one physical sciences and engineering laboratories corresponded with twenty-nine of the forty-two considered in the study made by Apstein in 1963. Many of these have changed somewhat in size, scope and function during the six-year period between the two surveys, but generally they were considered comparable.

The twenty-nine laboratories ranged in rank from first to almost last in both surveys. They were re-ordered to form a continuous ranking from first to twenty-ninth and were rank-order compared. The rank-order coefficient of correlation for the two lists was .86. The three largest differences were (1) the laboratory ranked eighth in Apstein's survey was now ranked twenty-third; (2) the laboratory ranked tenth was now ranked eighteenth; and (3) the laboratory ranked nineteenth was now ranked twenty-eighth.[1]

If the laboratories are again put in rank order without these three, the coefficient of rank-order correlation changes to .95. The rankings and ratings of the twenty-six laboratories are shown in Figure 2.17. The ratings from the 1963 survey

[1] At the time of the second survey, it had been decided to close two of the three omitted laboratories, although this was presumably not known to the majority of the raters. Whether or not the planned action influenced the changes in rank, or whether they would have been so ranked anyway, is unknown. In any event, the survey itself was not a factor in their subsequent closures; those actions had already been decided prior to the conduct of the survey.

1963 Survey		1969 Survey		
Rank	Rating	Rank	Rating	Rank-Order Difference
1	8.45	1	8.25	0
2	8.28	2	7.54	0
3	8.00	3	7.40	0
4	7.27	7	5.86	−3
5	7.16	6	6.09	−1
6	6.90	4	6.95	2
7	6.87	5	6.38	2
8	6.07	8	5.81	0
9	5.38	10	5.74	−1
10	5.34	9	5.79	1
11	5.33	12	4.86	−1
12	5.23	16	4.51	−4
13	4.91	18	4.32	−5
14	4.60	13	4.79	1
15	4.37	11	5.59	4
16	4.23	17	4.48	−1
17	3.98	14	4.73	3
18	3.93	15	4.69	3
19	3.90	22	3.90	−3
20	3.39	20	4.14	0
21	3.10	24	3.06	−3
22	3.08	19	4.22	3
23	3.00	25	3.05	−2
24	2.75	21	4.00	3
25	2.58	23	3.55	2
26	2.16	26	2.56	0

FIGURE 2.17
Comparison of Rankings From the 1963 Survey With Those From the 1969 Survey

span a wider range than those of the 1969 survey, probably because of the smaller 1963 sample size.[1,2]

One conclusion that may be drawn from the relatively high agreement between the two surveys is that they are both measuring the same thing, and doing this with a high degree of reliability (in the sense of repeatability). Another conclusion is that a laboratory's image changes quite slowly; except for the three that dropped radically, the relative order of the laboratories changed very little.

---

[1]The Apstein ratings were computed using the rankings of forty-two raters, whereas the 1969 ratings are based on the rankings of 280 raters.

[2]I do not know how many of the raters of the 1963 survey also participated in the 1969 survey; but even if they all did, and voted the same as before, this would contribute in only a minor way to the similarity between the ratings.

3. LABORATORY PROPERTIES DATA BASE

3.1 History, Purpose and Contents

For the past several years, the Office of Laboratory Management has annually collected and published data describing various quantitative properties of the DoD laboratories and other DoD RDT&E activities. In 1967 the procedure was formalized and expanded to form a data base containing information on staffing, funding, facilities, etc. Included also is information on individual missions, current important programs, functions, and facility capabilities.

In addition to providing a quick reaction capability to respond to the Congress and to other Agencies on very short notice and in many areas, this DoD-wide data base has been used by the Headquarters organizations of the three military departments to assist them in making management decisions. It has also served as a means of providing comparative information to Commanders and Laboratory Directors for assessment of their organizations with respect to others in their own Department and to laboratories in the other Services.

Data Elements

The data base elements fall into five general categories: staffing, facilities. appropriations, source of funding, and professional activity. The latter contains information on graduate training, publication of research, and attendance at meetings. The various categories are depicted in Figure 3.1; the appropriations category has been further divided into RDT&E sub-categories. With a few clues, the code names of the elements will be hopefully intelligible. For example, personnel items prefixed by MIL refer to military, by CIV to civilian; appropriations prefixed by IH are in-house, by OH are out-of-house. Appropriations suffixed R&D are for research and development, PRO are for procurement, and O&M are for operations and maintenance. DEP indicates that the source of funds is from the laboratory's own service, OTH means from some other source within the Department of Defense, and NON means not from within DoD. The prefix T means that the element is the total of some other elements, except for the condensation for technicians (TECHS).

The elements of the data base are defined in detail in DoD Instruction 7700.9 (Appendix A). Some of the footnotes in this chapter are quoted directly from the instruction. Seventy basic elements, together with thirty-five expanded elements (linear combinations of the basic seventy), are described briefly in Figures 3.2-3.4. Many of the expanded variables are used in publishing the annual report of laboratory properties [4]. The others were formed by the author in the expectation

PERSONNEL

MILBS	CIVBS	TBACH
MILMS	CIVMS	TMAST
MILPH	CIVPH	TPHDS
MILND	CIVND	TOTND
MPROF	CPROF	TPROF
MILSV	CIVSV	
MILST		
WGBRD	CLASS	TECHS
TAMIL	TACIV	TAPER

TRAINING AND PROFESSIONAL ACTIVITY

MILGS	CIVGS
MFTGS	CFTGS
PATNT	PAPER
RPRTS	MEETS

APPROPRIATIONS

IHR&D	OHR&D	TR&D$
IHPRO	OHPRO	TPRO$
IHO&M	OHO&M	TO&M$
IHOMA	OHOMA	
MILCN		
MILPA		
HOUSE		
TIH$	TOH$	TPGM$

SOURCE OF FUNDS

DEPRD	OTHRD	NONRD
DEPPR	OTHPR	NONPR
DEPOM	OTHOM	
DEPMS	OTHMS	NONMS
TDEP$	TODOD	TNDOD

RDT&E CATEGORIES

IH6.1	OH6.1	T6.1$
IH6.2	OH6.2	T6.2$
IH6.3	OH6.3	T6.3$
IH6.4	OH6.4	T6.4$
IH6.5	OH6.5	T6.5$
IH6.6	OH6.6	T6.6$
IHMPE	OHMPE	

LAND, PLANT, & EQUIPMENT

OWNED	LEASD	ACRES
RPROP	EQUIP	SEQIP
SEQPR	SEQNP	SEQAS
LSPAC	ASPAC	OSPAC

FIGURE 3.1

LABORATORY PROPERTIES DATA BASE

INDEX CODE

3	TAMIL	=	TOTAL AUTHORIZED MILITARY PERSONNEL
4	TACIV	=	TOTAL AUTHORIZED CIVILIAN PERSONNEL
5	MILBS	=	MILITARY WITH BACHELORS DEGREE
6	CIVBS	=	CIVILIANS WITH BACHELORS DEGREE
7	MILMS	=	MILITARY WITH MASTERS DEGREE
8	CIVMS	=	CIVILIANS WITH MASTERS DEGREE
9	MILPH	=	MILITARY WITH DOCTORS DEGREE
10	CIVPH	=	CIVILIANS WITH DOCTORS DEGREE
11	MILND	=	PROFESSIONAL MILITARY, NO COLLEGE DEGREE
12	CIVND	=	PROFESSIONAL CIVILIANS, NO COLLEGE DEGREE
13	WGBRD	=	NUMBER OF WAGEBOARD EMPLOYEES
14	CLASS	=	NUMBER OF CLASSIFIED ACT EMPLOYEES
15	TECHS	=	TECHNICIANS
16	CIVSV	=	PROFESSIONAL CIVILIAN SUPERVISORS
17	MILSV	=	PROFESSIONAL MILITARY SUPERVISORS
18	MILST	=	MILITARY SKILLED TRADES
19	OWNED	=	ACRES OWNED
20	LEASD	=	ACRES LEASED
21	RPROP	=	ACQUISITION COST OF REAL PROPERTY
22	EQUIP	=	ACQUISITION COST OF EQUIPMENT
23	LSPAC	=	LABORATORY SPACE
24	ASPAC	=	ADMINISTRATIVE SPACE
25	OSPAC	=	OTHER SPACE
27	SEQIP	=	SCIENTIFIC AND ENGINEERING EQUIPMENT
28	SEQNP	=	COST OF SEQIP, PRIOR FY NON-PROJECT MONEY
29	SEQPR	=	COST OF SEQIP, PRIOR FY PROJECT MONEY
31	IHR+D	=	RDT+E APPROPRIATIONS IN-HOUSE
32	OHR+D	=	RDT+E APPROPRIATIONS OUT-OF-HOUSE
33	IHPRO	=	PROCUREMENT APPROPRIATIONS IN-HOUSE
34	OHPRO	=	PROCUREMENT APPROPRIATIONS OUT-OF-HOUSE
35	IHO+M	=	OPERATIONS AND MAINTENANCE, IN-HOUSE
36	OHO+M	=	OPERATIONS AND MAINTENANCE, OUT-OF-HOUSE
37	MILCN	=	MILITARY CONSTRUCTION APPROPRIATIONS
38	MILPA	=	MILITARY PAY AND ALLOWANCES
39	IHOMA	=	OTHER APPROPRIATIONS IN-HOUSE

FIGURE 3.2

LABØRATØRY PRØPERTIES DATA BASE

INDEX CØDE

40	ØHØMA	=	ØTHER APPRØPRIATIØNS ØUT-ØF-HØUSE
41	HØUSE	=	HØUSEKEEPING AND ADMINISTRATIVE EXPENSES
42	IH6.1	=	IN-HØUSE RESEARCH DØLLARS
43	IH6.2	=	EXPLØRATØRY DEVELØPMENT
44	IH6.3	=	ADVANCED DEVELØPMENT
45	IH6.4	=	ENGINEERING DEVELØPMENT
46	IH6.5	=	MANAGEMENT SUPPØRT
47	IH6.6	=	ØPERATIØNAL SYSTEMS SUPPØRT
48	IHMPE	=	ALL ØTHER RDT+E
49	ØH6.1	=	ØUT-ØF-HØUSE RESEARCH DØLLARS
50	ØH6.2	=	EXPLØRATØRY DEVELØPMENT
51	ØH6.3	=	ADVANCED DEVELØPMENT
52	ØH6.4	=	ENGINEERING DEVELØPMENT
53	ØH6.5	=	MANAGEMENT SUPPØRT
54	ØH6.6	=	ØPERATIØNAL SYSTEMS SUPPØRT
55	ØHMPE	=	ALL ØTHER RDT+E
56	DEPRD	=	RDT+E FUNDS, DEPARTMENTAL
57	ØTHRD	=	RDT+E FUNDS, ØTHER DØD
58	NØNRD	=	RDT+E FUNDS, NØN-DØD
59	DEPPR	=	PRØCUREMENT FUNDS, DEPARTMENTAL
60	ØTHPR	=	PRØCUREMENT FUNDS, ØTHER DØD
61	NØNPR	=	PRØCUREMENT FUNDS, NØN-DØD
62	DEPØM	=	Ø+M FUNDS, DEPARTMENTAL
63	ØTHØM	=	Ø+M FUNDS, ØTHER DØD
64	DEPMS	=	ALL ØTHER FUNDS, DEPARTMENTAL
65	ØTHMS	=	ALL ØTHER FUNDS, ØTHER DØD
66	NØNMS	=	ALL ØTHER FUNDS, NØN-DØD
67	PATNT	=	PATENT APPLICATIØNS
68	PAPER	=	PAPERS PUBLISHED
69	RPRTS	=	TECHNICAL REPØRTS
70	CIVGS	=	CIVILIAN GRADUATE STUDENTS
71	MILGS	=	MILITARY GRADUATE STUDENTS
72	CFTGS	=	CIVILIAN FULL-TIME GRADUATE STUDENTS
73	MFTGS	=	MILITARY FULL-TIME GRADUATE STUDENTS
74	MEETS	=	TECHNICAL SØCIETY MEETINGS ATTENDED

FIGURE 3.3

LABORATORY PROPERTIES DATA BASE

INDEX CODE

75	MPRØF	= MILITARY PRØFESSIØNALS	(5+7+9+11)
76	CPRØF	= CIVILIAN PRØFESSIØNALS	(6+8+10+12)
77	TPRØF	= TØTAL PRØFESSIØNALS	(75+76)
78	TBACH	= TØTAL BACHELØRS	(5+6)
79	TMAST	= TØTAL MASTERS	(7+8)
80	TPHDS	= TØTAL PHDS	(9+10)
81	TAPER	= TØTAL AUTHØRIZED PERSØNNEL	(3+4)
82	TSPAC	= TØTAL SPACE	(23+24+25)
83	TR+D$	= TØTAL RDT+E	(31+32)
84	TPRØ$	= TØTAL PRØCUREMENT	(33+34)
85	TØ+M$	= TØTAL ØPERATIØNS AND MAINTENANCE	(35+36)
86	TPGM$	= TØTAL PRØGRAM DØLLARS	(SUM ØF 31-40)
87	TIH$	= TØTAL DØLLARS SPENT IN-HØUSE	
88	IH1-2	= IH6.1+IH6.2	
89	IH1-3	= IH6.1+IH6.2+IH6.3	
90	IH1-4	= IH6.1+IH6.2+IH6.3+IH6.4	
91	ØH1-2	= ØH6.1+ØH6.2	
92	ØH1-3	= ØH6.1+ØH6.2+ØH6.3	
93	ØH1-4	= ØH6.1+ØH6.2+ØH6.3+ØH6.4	
94	TDEP$	= TØTAL DØLLARS FRØM ØWN SERVICE	
95	TØDØD	= TØTAL DØLLARS FRØM ØTHER DØD	
96	TNDØD	= TØTAL DØLLARS FRØM ØUTSIDE DØD	
97	T6.1$	= TØTAL 6.1 DØLLARS	
98	T6.2$	= TØTAL 6.2 DØLLARS	
99	T6.3$	= TØTAL 6.3 DØLLARS	
100	T6.4$	= TØTAL 6.4 DØLLARS	
101	T6.5$	= TØTAL 6.5 DØLLARS	
102	T6.6$	= TØTAL 6.6 DØLLARS	
103	T61-2	= ALL 6.1 AND 6.2 DØLLARS	
104	T61-3	= ALL 6.1,6.2, AND 6.3 DØLLARS	
105	T61-4	= ALL 6.1,6.2,6.3, AND 6.4 DØLLARS	
106	TØH$	= TØTAL DØLLARS SPENT ØUT-ØF-HØUSE	
107	ACRES	= LAND ØWNED + LAND LEASED	
108	SEQAS	= SCIENTIFIC EQUIPMENT ACQUISITIØN	(28+29)
109	TØTND	= TØTAL PRØFESSIØNALS, NØ DEGREE	(11+12)

FIGURE 3.4

that they might prove more useful and/or illuminating than the basic elements. The gaps in the numbering of the elements correspond to non-numeric information that was at first carried over from the original data but was subsequently discarded.

The personnel elements (numbers 3-18) give information about the levels of various categories of staffing: military/ civilian, professionals/non-professionals, wageboard/classified, etc. The first two elements, Total Authorized Civilians (TACIV) and Total Authorized Military (TAMIL), represent ceilings as of the base-data of the report (end of the fiscal year); the remaining fourteen elements are on-board counts as of the base date.

The professional activity elements (numbers 67-74) contain data about the activity of people during the fiscal year being reported: patent applications, papers published,[1] technical reports,[2] training, etc. The categories of training are intended to be mutually exclusive, i.e., CIVGS includes civilians doing part-time graduate work and does not include those doing full-time study (CFTGS). MEETS is defined as the total number of people attending technical society meetings.

The facilities elements (numbers 19-29) contain data about plant, property, and equipment. LSPAC, ASPAC, and OSPAC refer to space used for laboratory, administrative, or other purposes. SEQIP is the acquisition cost of scientific and engineering equipment. SEQPR and SEQNP are the acquisition cost of scientific and engineering equipment obtained with project or non-project funds during the fiscal year being reported; SEQAS is an expanded variable, the sum of SEQPR and SEQNP.[3]

The remaining elements of the basic data have to do with money - the source of funds and type of appropriation. All of the financial data is reported in terms of "total obligational authority". This is defined as the total financial resources available for obligation in the specific year being reported on. This includes unobligated authority carried forward from the prior year and all obligational

---

[1] Papers published - Papers published must be original in-house work that have been published in an editorial review journal. It should not include state-of-the-art summaries or things of this nature. The author should be a full-time government employee at the time of writing.

[2] Technical reports - These are technical documentary reports related to scientific or engineering work and which are clearly identified by the laboratory as such. These must be prepared solely by in-house personnel as defined in footnote 1.

[3] There has been some inconsistency in the reporting of SEQIP and EQUIP. Some laboratories have reported as SEQIP only the acquisition cost of scientific equipment acquired during the reporting period; others have reported it as the total acquisition cost of all scientific and engineering equipment, but have subtracted it from EQUIP. It was (and is) intended that SEQIP represent the acquisition cost of all scientific and engineering equipment, and that the value reported in EQUIP include the value shown by SEQIP.

authority received or made available for obligation in the year being reported, including unobligated authority which will be carried forward into the subsequent year. The financial data is categorized according to type of appriation: Research and Development, Procurement, Operations and Maintenance, and Miscellaneous.

Elements 56-66 indicate the money in each appropriation category according to source of funding: Own Department, Other DoD, or Non-DoD. Elements 31-36 and 39-40 indicate the money in each category according to whether it is used In-House[1] or Out-of-House.[2] Elements 42-55 give a further break-down of the research and development dollars according to R&D appropriation sub-categories.

The financial data elements (elements 31-66) are reported in units of thousands of dollars, as are also elements 21-22 and 27-29 (having to do with real property and equipment). All other elements are expressed in natural units, except that in this report the space data (elements 23-25) are given in thousands of square feet.

3.2 Appropriations

Appropriations for each military department are authorized annually by Act of Congress [7]. They are generally intended to provide for expenses as indicated below, subject to the provisions and limitations of the individual Acts. RDT&E appropriations are for basic and applied scientific research, development, test, and evaluation, including the procurement of RDT&E supplies and materials and the rehabilitation, lease, and operation of facilities and equipment. Procurement appropriations are for the procurement, manufacture, and modification of missiles, armament, ammunition, vehicles, vessels, and aircraft, and for the acquisition, construction, and expansion of land, plant, and equipment. Operations and Maintenance appropriations are for expenses, not otherwise provided for, necessary for the operation and maintenance of the (specified military department), i.e., modification of missiles and ordnance; alteration of aircraft and vessels; exercises and

---

[1] In-House: Total obligational authority reported under this category are for activities performed or to be performed by the organizational entity. Their work is carried on directly by their own personnel. This item includes the costs of supplies and equipment, essentially of an "off the shelf" nature, which are procured for use in in-house research and development plus such things as travel, publications, and other types of services in support of in-house functions. Excluded from the in-house total are the expenses for planning and administering programs by DoD personnel, including military, of contracts and grants for out-of-house work.

[2] Out-of-House: Total obligational authority reported under this category are for activities performed or to be performed by other than the organizational entity. Out-of-house performers may include other departmental or DoD organizational entities, industrial firms, educational institutions, not-for-profit institutions and private individuals. Included as out-of-house work are all expenses paid the out-of-house performes as well as the expenses of the organizational entity. This also includes travel and other supporting services.

maneuvers; transportation of things; repair and maintenance of facilities; training and care of personnel; general administration, etc.

There is an RDT&E appropriation for each military department; the same is true for O&M appropriations. In procurement, there is one appropriation for the Army, and three each for the Navy and the Air Force. These latter are listed below, although elsewhere in this report they are considered as procurement in the aggregate sense.

- Procurement of Equipment and Missiles, Army

- Procurement of Aircraft and Missiles, Navy
- Shipbuilding and Conversion, Navy
- Other Procurement, Navy

- Aircraft Procurement, Air Force
- Missile Procurement, Air Force
- Other Procurement, Air Force

The average total appropriation to the fifty-one physical sciences and engineering laboratories during fiscal years 1967, 1968, and 1969 was 2.46 billion dollars. This was apportioned 71% to RDT&E, 18% to Procurement, 4% to Operations and Maintenance, and 7% to Miscellaneous. Miscellaneous appropriations include military pay and allowances, military construction, and funds not specifically identified with the type of appropriation (RDT&E, O&M, or Procurement). These latter generally are associated with inter-laboratory and inter-departmental transfers of funds, and possibly are being counted twice: as Out-of-House Miscellaneous Appropriations by the issuing laboratory, and as In-House Other Miscellaneous Appropriations by the receiving laboratory. In a few instances they exceed ten percent of a laboratory's total appropriations, but for the most part they are of the order of five percent or less.

The distribution by military department (Figure 3.5) shows the variations by service: the RDT&E portion of the Air Force appropriation was substantially greater than the corresponding proportion in the Navy; the Navy on the other hand had a much larger percentage of procurement dollars. The Army proportions were almost identical to the average of services

1967-68-69 Average Annual Appropriations
Physical Sciences and Engineering Laboratories

	Army	Navy	Air Force	DoD
	Dollars in Millions			
R&D	682	588	471	1741
PRO	166	241	38	445
O&M	39	60	7	106
Misc.	74	72	21	167
TPGM	961	961	537	2459
	Percent Distribution			
R&D	71	62	88	71
PRO	17	25	7	18
O&M	4	6	1	4
Misc.	8	7	4	7

FIGURE 3.5

% Appropriations by Labs FY 67-68-69

	Army	Navy	Air Force
	23 Labs	18 Labs	10 Labs
R&D	76	64	91
PRO	13	20	5
O&M	5	7	1
Misc	6	9	3
	20 Labs	16 Labs	8 Labs
R&D	82	71	97
PRO	10	13	- -
O&M	4	7	- -
Misc	4	9	3

FIGURE 3.6

The upper section of Figure 3.6 shows the average of the distribution of the appropriations within each of the individual laboratories for the three military departments. The RDT&E portion tends to be somewhat higher when averaged by laboratories than when taken for the departments as a whole (compare to Figure 3.5). The lower section of Figure 3.6 shows the percent distribution by laboratories when the few most exceptional laboratories are omitted; the Air Force laboratories in this case are seen to be almost entirely funded by RDT&E appropriations; the Navy laboratories have a substantial portion in miscellaneous and O&M.

Figure 3.7 shows the general categories of appropriations for fiscal year 1968, with further information on the distribution of dollars in-house and out-of-house, and according to the R&D sub-category (6.1 to 6.6). The Army was fairly evenly distributed between in-house and out-of-house RDT&E; the Navy allocated more than 70% of its RDT&E appropriations in-house; the Air Force assigned more than 75% of its RDT&E appropriations out-of-house (see Figure 5.1).

3.3 Correlation Between Elements

In looking for associations between the peer ratings and sets of data elements, it would be desirable to know the correlation between data elements and to use those that are the least mutually correlated, since if one element is highly correlated with another, it cannot be expected to significantly augment the degree of association already established by the first. Even more, if two or more dependent elements are used to measure laboratory effectiveness, it is difficult to estimate the change in effectiveness attributable to a change in one of the elements if there is a substantial amount of correlation between them. A knowledge of the association between elements would also be useful in reducing the number of elements to be taken under consideration. While this has not been an aim of the present study - in which the number of elements has been increased through the addition of the expanded variables - such a reduction of variables would prove useful in future examination.

A measure of the inter-relationships between the data elements can be obtained by examining the correlations between variables, and grouping together those that are significantly correlated. Among the basic elements of the data for fiscal year 1968, there were 117 pairs with a correlation of .70 or greater; these are listed in Figure 3.8. Some of the groups thus identified from these pairs of data are shown in Figure 3.9.

Systematic procedures exist for identifying groups of related variables. Two such methods are cluster analysis and factor analysis. In cluster analysis, a variable is assigned to a cluster to which it seems most to "belong", according to an

Appropriations to DoD Laboratories
(In Millions of Dollars)
Fiscal Year 1968

	Army	Navy	Air Force	DoD
R&D	682	614	418	1719
PRO	179	257	44	480
O&M	35	53	8	96
Misc.	26	75	23	124
TPGM	927	999	493	2419
In-House	434	603	116	1153
Out-of-House	493	396	377	1266
IHR&D	341	432	98	871
OHR&D	346	182	320	848
T6.1	51	54	66	171
T6.2	172	176	233	581
T6.3	135	109	59	303
T6.4	133	95	29	257
T6.5	46	38	10	94
T6.6	72	89	7	168
Other R&D	78	53	14	145

FIGURE 3.7

PAIRS ØF VARIABLES WITH CØRRELATIØNS GREATER THAN .7

MILMS,MIL8S .842	MILGS,MIL8S .807	MILGS,MILMS .893
TECHS,MILPA .732	TAMIL,MILPA .834	ØH6.3,MILSV .697
CIVBS,TACIV .893	CIVMS,TACIV .797	WG8RD,TACIV .885
CLASS,TACIV .980	TECHS,TACIV .847	CIVSV,TACIV .787
LSPAC,TACIV .720	IHR+D,TACIV .863	IHPRØ,TACIV .750
HØUSE,TACIV .864	IH6.3,TACIV .701	CIVMS,CIVBS .826
CLASS,CIVBS .921	TECHS,CIVBS .769	CIVSV,CIVBS .877
IHR+D,CIVBS .899	HØUSE,CIVBS .770	IH6,3,CIVBS .734
IH6.4,CIVBS .787	DEPRD,CIVBS .789	PATNT,CIVBS .727
CIVPH,CIVMS .747	CLASS,CIVMS .848	TECHS,CIVMS .784
CIVSV,CIVMS .770	IHR+D,CIVMS .880	IH6.3,CIVMS .743
DEPRD,CIVMS .737	IH6.1,CIVPH .850	PAPER,CIVPH .808
TECHS,CLASS .881	CIVSV,CLASS .841	LSPAC,CLASS .724
IHR+D,CLASS .904	HØUSE,CLASS .884	IH6.3,CLASS .756
DEPRD,CLASS .716	IHR+D,CIVSV .781	IH6.3,CIVSV .724
IH6.4,CIVSV .759	DEPRD,CIVSV .768	IH6.3,CIVND .730
IH6.5,CIVND .731	ØH6.3,CIVND .736	DEPRD,CIVND .743
ØWNED,TECHS .734	LSPAC,TECHS .737	IHR+D,TECHS .865
IH6.3,TECHS .772	HØUSE,TECHS .702	DEPRD,TECHS .734
CIVSV,TECHS .734	CLASS,WGBRD .776	RPRØP,WGBRD .703
IHPRØ,WGBRD .812	HØUSE,WGBRD .774	LEASD,ØWNED .914
RPRØP,ØWNED .856	ØSPAC,ØWNED .934	ØTHØM,ØWNED .977
ØTHMS,ØWNED .906	SEQNP,ØWNED .849	RPRØP,LEASD .769
ØSPAC,LEASD .859	SEQNP,LEASD .807	ØTHØM,LEASD .963
ØTHMS,LEASD .874	ØSPAC,RPRØP .917	ØTHØM,RPRØP .830
ØTHMS,RPRØP .770	SEQNP,RPRØP .830	HØUSE,RPRØP .749
SEQIP,EQUIP .721	LSPAC,EQUIP .716	IHR+D,EQUIP .711
IH6.2,EQUIP .744	DEPRD,EQUIP .728	IHR+D,LSPAC .732
IH6.2,LSPAC .747	HØUSE,IHR+D .752	IH6.2,IHR+D .796
IH6.3,IHR+D .809	IH6.4,IHR+D .765	DEPRD,IHR+D .823
SEQNP,ØSPAC .896	IHPRØ,ØSPAC .745	HØUSE,ØSPAC .771
ØTHØM,ØSPAC .914	ØTHMS,ØSPAC .843	HØUSE,SEQNP .762
ØTHØM,SEQNP .843	ØTHMS,SEQNP .793	IHPRØ,SEQNP .726
ØTHMS,DEPØM .887	DEPØM,IHØ+M .858	HØUSE,IHPRØ .729
DEPPR,ØHPRØ .963	DEPMS,IHØMA .697	TECHS,EQUIP .694
PATNT,IHR+D .728	PAPER,IH6.1 .735	MEETS,IH6.1 .704
ØH6.3,IH6.3 .717	DEPRD,IH6.3 .794	MEETS,PAPER .740
IH6.6,IH6.4 .705	ØH6.6,IH6.4 .726	DEPRD,IH6.4 .711
ØH6.6,IH6.6 .750	PATNT,IH6.6 .697	ØH6.3,ØHR+D .846
DEPRD,ØHR+D .889	DEPRD,ØH6.3 .826	DEPRD,ØH6.4 .762
PATNT,CIVMS .695	ØTHRD,ØHMPE .842	IH6.2,CIVBS .696
TECHS,IH6.2 .692	CLASS,IH6.4 .698	CLASS,IH6.2 .693

FIGURE 3.8

Correlation Between Variables

	CIVBS	CIVMS	CLASS	TECHS	CIVSV	IHR&D
TACIV	.89	.79	.98	.84	.78	.86
CIVBS		.82	.92	.76	.87	.89
CIVMS			.84	.78	.77	.88
CLASS				.88	.84	.90
TECHS					.73	.86
CIVSV						.78

	LEASD	RPROP	OSPAC	SEQNP
OWNED	.91	.85	.93	.84
LEASD		.76	.85	.81
RPROP			.91	.83
OSPAC				.89

	IH6.1	PAPER	MEETS
CIVPH	.85	.81	.70
IH6.1		.73	.70
PAPER			.74

FIGURE 3.9
Selected Groups of Correlated Variables
FY 68 Data Base

algorithmic procedure. After variables have been assigned to clusters, it is incumbent upon the researcher to identify or categorize thc nature of the cluster.

Factor analysis is similar to cluster analysis, except that in cluster analysis it is customary to assign a variable entirely to one cluster, whereas in factor analysis a variable is usually divided into portions which are assigned to a number of different factors. As with cluster analysis, it is up to the investigator to determine the nature of the factor; the method simply indicates what proportions of a variable contributes to the particular factor.

The data for fiscal year 1968 were factor-analyzed using a program obtained from the Biometric Laboratory of the University of Miami. Groupings were obtained for the fifty-one laboratories collectively and also by individual military departments; the results are shown in Appendix B. The three groups previously shown in Figure 3.9 were among the principal groupings identified; others included a coupling of military professionals with out-of-house dollars; civilian professionals (no degree) with in-house 6.3 dollars; and operations and maintenance with procurement appropriations. For the most part, however, the groupings are different for the separate services, and for DoD as a whole.

The factor analyses were not undertaken until late in the study, when other major aspects of the study were either completed or well underway. Consequently they were not used except in a supplementary and corroborative sense, nor has any more than a cursory attempt been made to interpret the various factors.[1] I do not know how useful they would have been had they been available at the outset; I presume they would have provided insight into relationships between the variables which were otherwise only learned as by-products of other activity.

3.4 Distributional Characteristics of Laboratory Properties

In considering the elements of the laboratory properties data base, particularly with respect to the selection of variables to be used in the investigation of associations between the peer ratings and the data base, various questions arise having to do with the variation and distribution of the elements. Is a particular property representative of laboratories in general, or of one or two laboratories in particular? How are the various elements distributed among the laboratories? How do the data elements vary from year to year? In addressing these and similar questions,

---

[1] See "*Derivation of Theory by Means of Factor Analysis, or Tom Swift and His Electric Factor Analysis Machine*," by J. S. Armstrong, **The American Statistician**, December 1967, for a penetrating review of the application (and mis-application) of factor analysis.

it has been convenient to examine three attributes of the distribution of laboratory properties: density, dispersion, and stability. Density is a measure of the extent to which a property is common to the population of laboratories; dispersion refers to the distributional variation of the elements; and stability is concerned with the variation of the elements over time, i.e., the amount of annual change. Summary information pertaining to these attributes is presented in Appendix C; where appropriate, the data is also presented in the present section.

Density

By density is meant the proportion of laboratories that have non-zero values of a particular element. For example, if most of the laboratories have some non-zero portion of an element, that element is considered dense; if only a few laboratories have the characteristic, it is considered sparse. In general, the denser elements are more representative than those that are sparse, although not necessarily more informative. More important, if two elements have the same correlation with the laboratory ratings, the one that is denser will be the more statistically significant.

The distribution of a property within a military department will be said to be "dense" if the element is shared by at least a certain upper proportion of the department's laboratories, and "sparse" if it is common to no more than a certain lower proportion. Applying a proportion of one-sixth to both the upper and lower parts of the sixty-nine basic elements of the data for fiscal year 1968 yields the following set of numbers:

	Army	Navy	Air Force
Dense	32	46	35
Sparse	4	3	7

whereas a choice of one-third for the upper and lower proportions gives the numbers shown below.

	Army	Navy	Air Force
Dense	49	56	37
Sparse	9	5	15

i.e., there are forty-nine elements that are common to at least sixteen Army laboratories, and nine elements that are *not* common to more than eight Army laboratories.

The number of elements that are completely dense, i.e., possessed in some non-zero amount by all of the laboratories within a department is presented below; the second line shows the number of elements that are common to all but one of the laboratories.

	Army	Navy	Air Force
All Laboratories	16	34	26
All But One	22	42	35

Considering the two tables together, it can be seen that the properties of the Navy laboratories are substantially more dense than are those of the other two services; and that those of the Army and the Air Force are approximately of the same relative density.

Overall, the elements that are most dense are the staffing elements (except for Military Professionals No Degree); the plant and facilities-elements (except for Land Leased); some of the RDT&E elements (IHR&D, OHR&D, IH6.1, IH6.2, OH6.2 DEPRD); and the training and productivity elements (except for Military Graduate Students). The elements that are most sparse are Military Professionals No Degree (MILND), Land Leased (LEASD), Military Construction (MILCON), Out-of-House Miscellaneous Appropriations (OHOMA), Out-of-House Operational Systems Support (OH6.6), Military Full-Time Graduate Students (MFTGS), and most of the source of funding from outside the parent military department (OTHPR, OTHOM, OTHMS, NONRD, NONPR,[1] NONMS).

The thirty-five expanded elements were mainly in the upper part of the density profile. There was one marginal entry - T6.6$ - for all three. The Air Force had three additional marginal or low-order elements:

T6.4$	(5 out of 10)
TPRO$	(3 out of 10)
TO&M$	(5 out of 10)

Dispersion

A number of descriptors were considered in trying to find one that would serve as a topic heading for this sub-section. Some of these were: uniformity (non-uniformity); symmetry (asymmetry); parity (disparity); and regularity (irregularity). Dispersion was finally selected as a label representative of these various characteristics.

[1]This element is entirely sparse for all services, all years.

Dispersion in this sense refers to the way an element is distributed among the different laboratories. At one extreme, the laboratories could have an equal or nearly equal amount of an element, in which case the element would have little power for resolution or discrimination between laboratories. At the other extreme, one laboratory may have several times as much as any one of the others; this value may be an extreme point, lying several standard deviations away from the mean. Even one such outlier can mask associations that might otherwise be seen, or can indicate a strong association where such is not really the case.

The principal attributes that characterize the distribution of the data base elements among the various laboratories are the asymmetry of the frequencies and the disparity of the magnitudes. For almost all elements, a few laboratories account for a large part of the total; and in almost all cases the majority of the laboratories are clustered in the lower half of the distribution.

For example, the distribution of some of the principal elements of the fifty laboratories are shown in Figure 3.10. These represent the average value of the elements over the three fiscal years 1967, 1968, and 1969. The distribution has been apportioned into ten equal intervals by dividing the difference between the maximum and minimum values of the elements by ten. Most of the laboratories are clustered in the lower (left-hand) part of the distribution.

	NUMBER OF LABORATORIES PER INTERVAL									
Total Professionals	11	11	10	8	4	1	2	1	-	1
Total Personnel	16	12	11	3	4	1	1	1	-	1
Total R&D Dollars	23	15	3	4	3	1	-	-	-	1
Total Program Dollars	15	13	8	1	9	2	-	-	-	2

FIGURE 3.10
Distribution of Laboratory Properties

The five laboratories in the upper part of the distribution of professionals account for more than 25% of the total. In the distribution of the R&D program element, 96% of the laboratories are in the lower five intervals; 25% of the laboratories (the upper eight intervals) account for about 50% of the total DoD laboratory research program.

This latter proportion is by no means atypical; in many cases twenty to twenty-five percent of the laboratories having a particular property account for more than fifty percent of its total weight, and generally, less than ten percent of the

laboratories account for more than twenty-five percent of each element. In all cases, it takes no more than five laboratories to account for one-quarter of an element's value, and usually, less than five. The numbers of laboratories making up 25%, 50%, 75%, and 100% of the value of each element are shown in Table 1 of Appendix C; these have been computed using the fiscal year 1968 data for the fifty-one physical sciences and engineering laboratories. The cumulative quartile distributions of some of the more representative elements are shown below.[1]

	NUMBER OF LABORATORIES ACCOUNTING FOR			
	25%	50%	75%	100%
Total Professionals	5	13	25	51
Total Personnel	5	12	24	51
Total RDT&E Program	4	10	22	51
Total Laboratory Program	5	11	22	51

The same pattern generally applies to the laboratories within the military department. For example, using the same data elements as above, also for fiscal year 1968, the cumulative quartile distributions within the Navy laboratories were

	NUMBER OF LABORATORIES ACCOUNTING FOR			
	25%	50%	75%	100%
Total Professionals	3	6	10	18
Total Personnel	2	5	10	18
Total RDT&E Program	2	4	9	18
Total Laboratory Program	2	5	10	18

[1] Alternatively, the boundaries could have been selected to represent the quartile points of the number of laboratories, in which case the data would have been presented in the form

	PERCENT OF TOTAL ACCOUNTED FOR BY FOLLOWING NUMBER OF LABORATORIES			
	13	26	39	51
Total Professionals	50%	76%	92%	100%

from which can be derived the frequency distribution by quartiles.

	NUMBER OF LABORATORIES IN QUARTILE			
	I	II	III	IV
Total Professionals	3	3	4	8
Total Personnel	2	3	5	8
Total RDT&E Program	2	2	5	9
Total Laboratory Program	2	3	5	8

Two of the Navy laboratories receive at least 25% of the funds allocated to all eighteen, three more account for the next 25%, then five; and finally eight.[1,2]

A technique for illustrating the tendency of a few laboratories to account for a large part of the distribution of an element is the Lorenz curve. This is shown graphically in Figure 3.11 for the eighteen Navy laboratories. Chart (1) contains four reference curves. Curve a represents the line of even distribution, i.e., all laboratories having an equal part of the whole. Curve b shows how the distribution would look if the elements were proportional to the sequence of natural numbers 1, 2, 3, ···, 18. Curve c is drawn assuming the elements are distributed according to a Fibonnaci sequence. Curve d portrays a distribution based on the binary sequence 1, 2, 4, 8, ··· . Chart (2) diagrams the distribution of total professionals using the 1968 data; Chart (3) shows the distribution of the total R&D program; Chart (4) illustrates the distribution of procurement dollars.

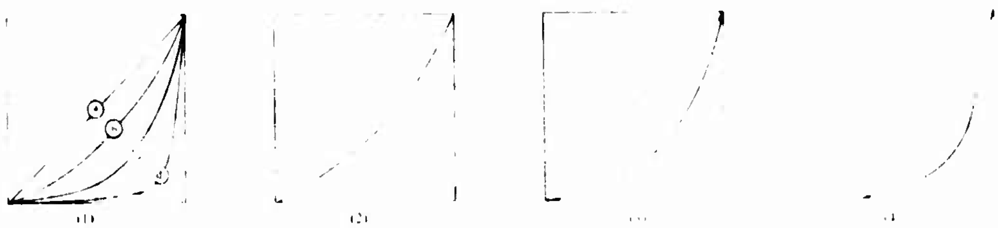


FIGURE 3.11
Cumulative Distributions of Selected Laboratory Properties

[1] I note in passing that the numbers 2,3,5,8 are part of the Fibonacci sequence 1,1,2,3,5,8,13,21,··· wherein each number is the sum of the previous two. The sequence frequently has application to phenomena of natural growth (e.g., spirals in sunflowers, pine cones, shells, etc.).

[2] This pattern is more typical of the Navy laboratories than of those of the other two services. The Air Force distribution of professionals is 2-2-2-4; the Army laboratories show 2-3-5-13, the fourth quartile being larger than the sum of the other three because of the larger proportion of smaller laboratories in the Army.

The purpose in presenting the distributional information in the foregoing way was to accentuate the effect of asymmetries in the frequencies and disparities in the magnitudes. In all four cases shown in Figure 3.10, the ratio of the maximum to the minimum values is sufficiently large that the class intervals can effectively be regarded as normalized intervals, i.e., the numbers 1, 2, 3, ···, 10 can be used as scaled approximations of the actual magnitudes. Thus for total professionals, eleven laboratories have unit value "1", ten have unit value "2", etc. This is generally true of most of the laboratory properties in the data base.

Another way of viewing these aspects of the distribution of laboratory properties is to compute the mean, the standard deviation, and the higher moments. For example, using the interval values and frequencies shown in Figure 3.10, the distribution of total professionals has a computed mean of 3.5 and a standard deviation of 2.2; the uppermost value is just barely within three sigma units of the mean. The distribution of research dollars has a substantially smaller dispersion; its mean is 2.2 and the standard deviation is 1.7. Consequently its uppermost value is even more extreme, lying more than four sigma units from the mean.

Probably the best way to view the distribution of laboratory properties is to look at them in some aggregated way, as was shown in Figure 3.10. These were derived from the distributions of the average values of the data base elements for fiscal years 1967, 1968, and 1969, which are shown in Table 6 of Appendix C; the range has been partitioned into fifty intervals. (These are the distributions for only those laboratories which had non-zero values of the elements. However, the disparity between the highest and lowest values in all cases is so large that the distribution including zeros looks practically the same as that without zeros; the number of zero values can simply be added to the number in the first interval shown.)

The deci-distributions for the individual military departments are shown in Appendix E under the tables headed "SKIP ZEROS"; these were computed from the 1968 data for those laboratories which had non-zero values of the elements. The intervals are arranged from lowest to highest in left to right order. The number under the column labelled "M/M" shows the ratio of the maximum to the minimum values for each element. For example, the number 5E 0 indicates that the value of the ratio is 5×10^0, or 5; the number 2E 1 indicates a ratio of 20. As long as the ratio is close to ten or greater, the interval distribution is an adequate representation of the actual distribution.[1]

[1]The disparity by individual services is sometimes so small that the inclusion of zeros seriously alters the distribution. The distributions including zeros are therefore shown separately in Appendix E in the tables labelled "COUNT ZEROS". In these tables, whenever one or more of the laboratories has a zero value, the number under M/M shows the value of the largest element rather than the ratio of the highest to the lowest (since in this case the value of the lowest is zero).

The distributions of some of the principal elements within the three military departments are shown in Figures 3.12 and 3.13; these are based on the average of the data for fiscal years 1967, 1968, and 1969. In Figure 3.12, the spread of the number of total professionals in each of the services is indicated by the position on the horizontal axis; nine of the Army laboratories had less than two hundred professionals, whereas the Air Force and the Navy each had only one laboratory in this category.[1] The vertical dimension shows the magnitude of the in-house R&D program; this was selected in preference to the total R&D program as being more representative across the military department.[2] The oblique dotted line represents a ratio of $40,000 per professional; the vertical dotted line shows the scale of a forty million dollar in-house RDT&E program at a staffing level of one thousand professionals.

Figure 3.13 illustrates a similar relationship between the total number of authorized personnel and the total laboratory program. The horizontal and vertical scales are each three times those used in the preceding figure, hence the oblique dotted line again indicates an average funding of forty thousand dollars per person. The vertical dotted line depicts a one hundred and twenty million dollar program at a laboratory with three thousand people.

The nature and form of the distribution of laboratory properties suggests that a logarithmic transformation might give a more symmetrical distribution of the frequencies; the results of such a transformation upon the non-zero values of the elements are shown in Table 7 of Appendix C, again partitioned into fifty zones The interval distributions of the original and the transformed values of the principal staffing and funding variables are shown in Figure 3.14. The logarithmic distributions tend to be mound-shaped, centered somewhat to the right, whereas the untransformed values tend to accumulate on the extreme left.

The distributions of the transformed elements by individual military department are presented in Appendix J under the tables headed "SKIP ZEROS".[3] Notice that

---

[1] Professionals include both degree and non-degree personnel performing professional activity, as defined in DoDI 7700.9.

[2] Since 70% of the Navy R&D program is conducted in-house versus 30% for the Air Force, there would be proportionally greater change in the Air Force ratios than in the Navy ratios if the vertical dimension had been chosen to represent the total R&D program. The Navy magnitudes would be less than double those shown, the Army magnitudes would be about twice again as large and the Air Force magnitudes would be about four times as large, with two of the lower three being eighteen times as large.

[3] It was desired to apply the logarithmic transformation to all values of the elements for use in correlation and regression analyses. In order to do this, an integer "1" was added to the values before taking the logarithm. The results of this transformation are shown in Appendix E under the tables headed "COUNT ZEROS"; the distributions shown are those of LOG(X+1).

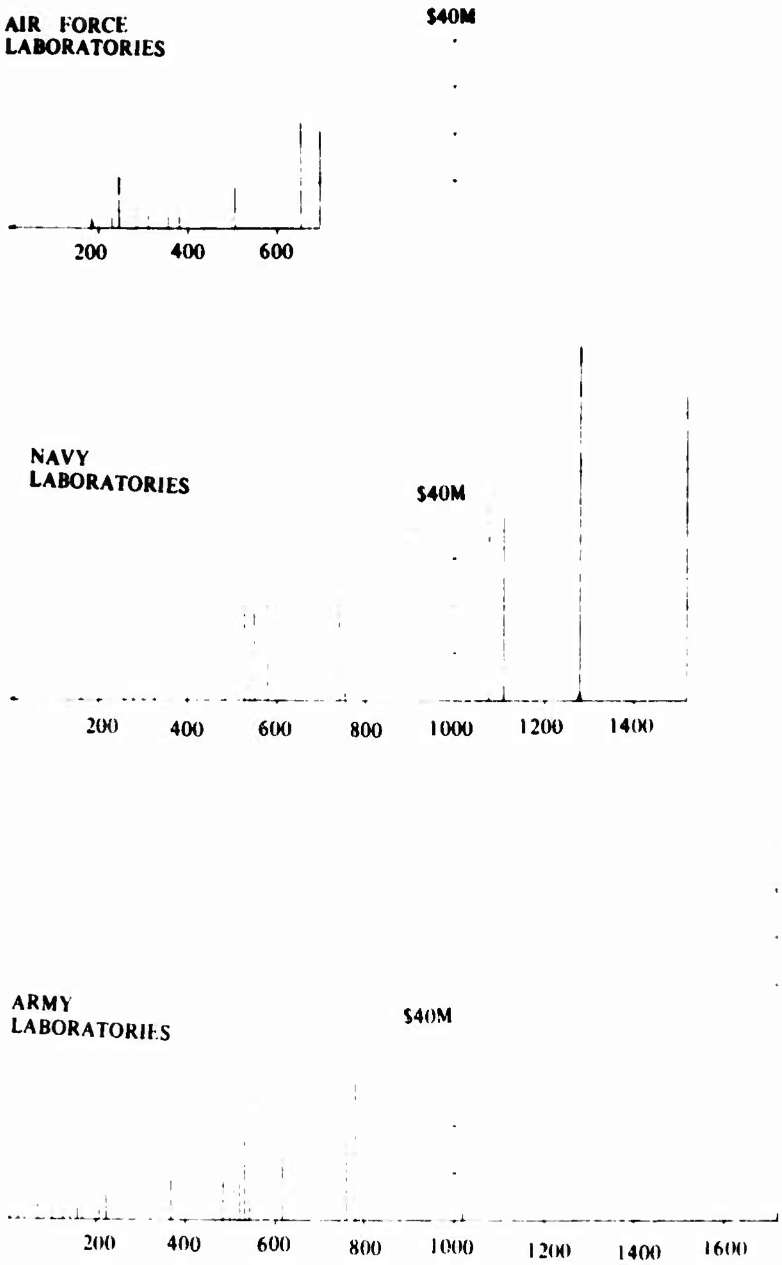


FIGURE 3.12
In-House RDT&E vs Number of Professionals, Average of FY 67-68-69 Data

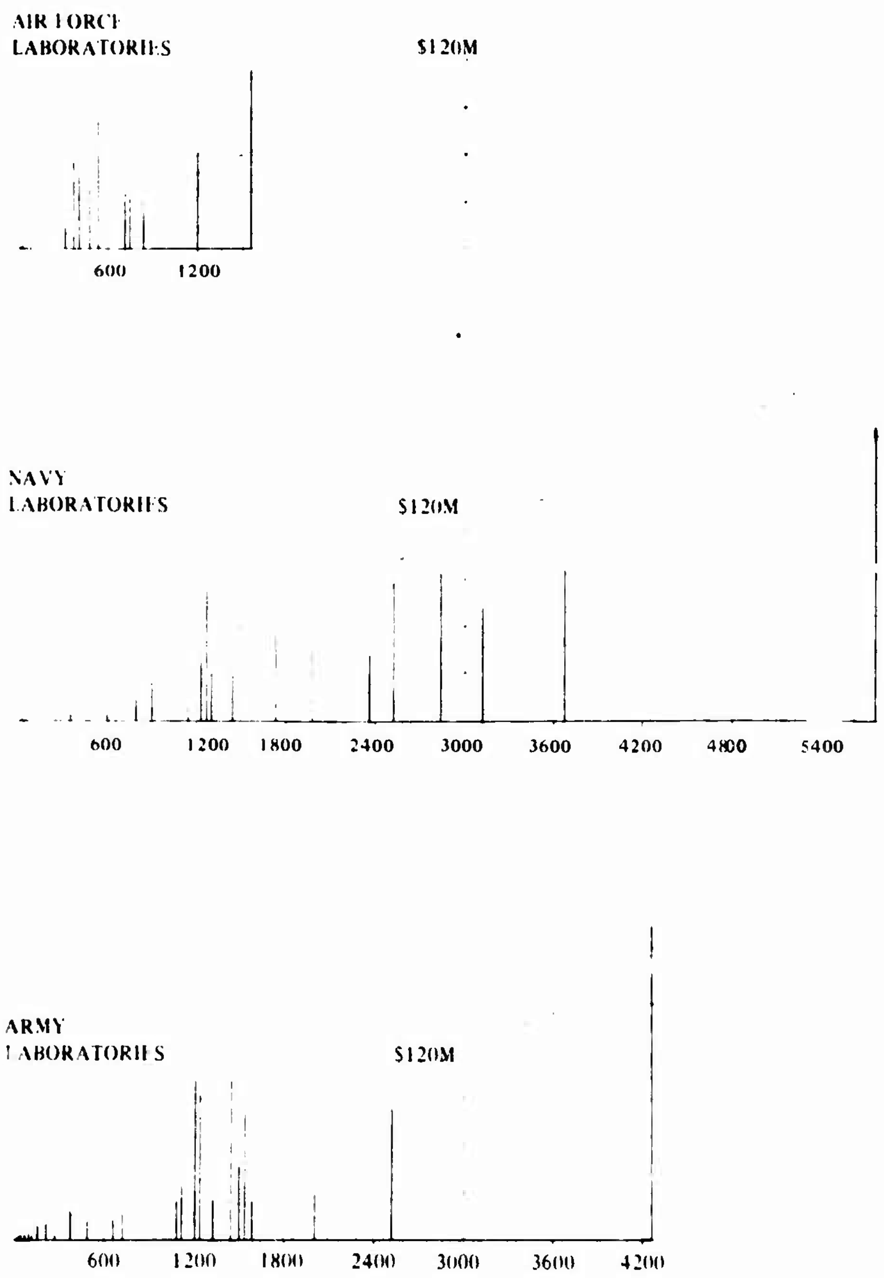
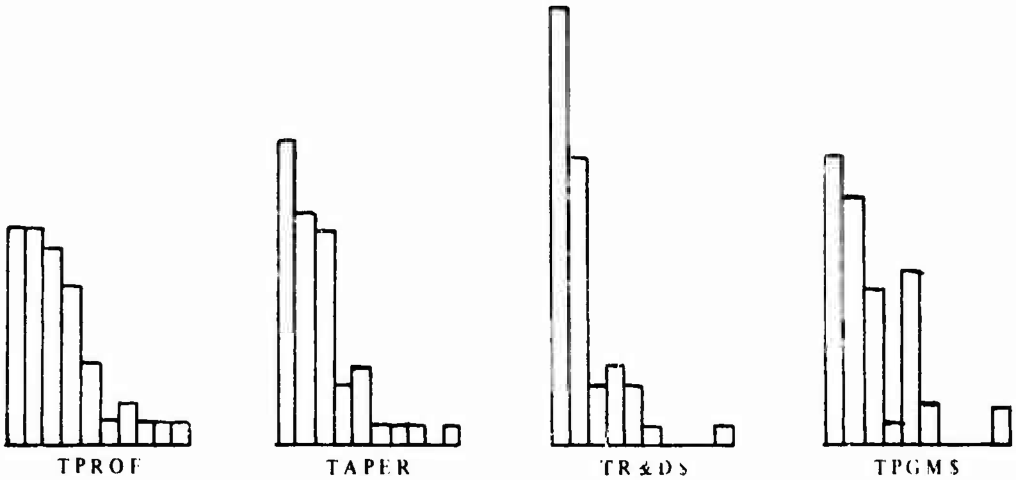
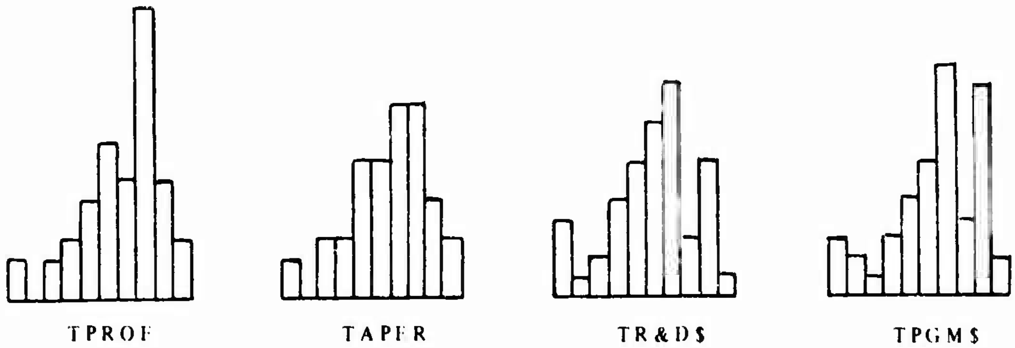


FIGURE 3.13
Total Program Dollars vs Total Authorized Personnel
Average of FY 67-68-69 Data



50 LABS - REGULAR



50 LABS - LOGS

FIGURE 3.14

the range of the variables is quite small, the largest being on the average only two or three times the size of the smallest. These transformation and other normalizations are discussed in Chapter 4.

Stability

Stability has to do with the variation of the laboratory properties with respect to time. Two ways have been used to measure the amount of variation in each of the elements; one utilizes the annual relative deviation from the mean of the three years of the data base; the other involves the relative change in the data between consecutive years. In both methods, the distributions are based on the maximum changes experienced by a laboratory over the three fiscal years. In the case of the mean deviation, the use of averages - either absolute values or sums of squares - would most likely show the same relative patterns of stability as obtained here; nor would I expect to see any significant change in the measurement of the relative stabilities by similarly averaging the annual fluctuations.

The maximum relative deviation is computed by dividing the absolute value of the maximum deviation from the mean by the value of the mean. For example, if the values of an element for a particular laboratory were 600, 500, and 100, the mean would be 400. The deviations from the mean would be -200, -100, and 300; the relative deviations would be 50%, 25%, and 75%; this latter would represent the maximum relative deviation from the mean. (If the mean of an element for a particular laboratory is zero, the laboratory is not included in the distribution of that element.)

Another way of measuring the stability of the data from year to year is in terms of percent annual change. Using the data from the above example, the value for one year would be five times that for an adjacent year. The change in the last two years might be regarded as an eighty percent decrease from year two to year three, or as a four hundred percent change relative to year three. In the table of percent annual deviation the variation is always taken relative to the smaller of two consecutive years (i.e., like the four hundred percent change in the example above). The larger of the two values is taken as the numerator; the ratio computed (the ratio is considered to be ∞ if the denominator is zero, except that 0/0 is not counted at all); the unit 1 is subtracted to determine the amount of change; and the result is expressed in percent.

The distribution of variation in the elements according to each of the two methods is shown in Tables 4 and 5 of Appendix C. The distributions of selected elements are shown in Tables A and B of Figure 3.16. The first line of Table A

indicates that nine of the twenty-three Army laboratories had changes of more than 50% from their mean value of Total Authorized Military personnel (TAMIL), whereas only one of the ten Air Force laboratories had a change in excess of 10%. The first line of Table B shows a similar situation, indicating that eight of the Army laboratories had at least one annual change in TAMIL greater than 50%, whereas nine of the ten Air Force laboratories had no annual change in excess of 20%.

The intervals in Table A were chosen to represent good stability (0% - 10%), medium stability (10% - 25%), poor stability (25% - 50%), and little or no stability (50% - 200%). (The definitions of good, medium, and poor were arbitrarily introduced by the author.) The number of laboratories having a maximum deviation between 0 and 10 percent is shown under the column labeled A; the number between 10 and 25 percent under the column head B; etc. The mathematics of the procedure are such that the maximum theoretical deviation is 200 percent.[1] The intervals in Table B were selected to roughly correspond to those in Table A.[2]

Using the table of percent deviation from the mean, and arbitrarily assuming that a definition of stability of an element is that at least two-thirds of the laboratories having that element have less than a 25% maximum deviation, the most stable elements for all three services are

TACIV	- Total Authorized Civilian Personnel
CIVBS	- Civilians with Bachelors Degree
CIVPH	- Civilians with PhD Degree
WGBRD	- Number of Wageboard Employees
CLASS	- Number of Classified Employees
OWNED	- Amount of Land Owned
RPROP	- Value of Real Property
CPROF	- Number of Civilian Professionals
TPROF	- Total Number of Professionals
TBACH	- Total Number of Bachelors
TMAST	- Total Number of Masters
TAPER	- Total Authorized Personnel
TSPAC	- Total Floor Space
IH1-2	- In-House Research and Exploratory Development
T61-2	- Total Research and Exploratory Development
ACRES	- Amount of Land Owned and Leased

[1] Although the relative deviation from the mean was used in preference to the relative standard deviation, the latter would have just about the same distribution as the former if the zone limits were divided by the square root of two.

[2] Alternatively, had the variations in Table B been obtained by using the larger of the two values as the denominator, the same distribution would be obtained if the zone limits were 0-17%, 17%-33%, 33%-50%, and 50%-100% (100% would imply a change to zero).

TABLE A

DISTRIBUTION OF DATA BASE ELEMENTS ACCORDING TO MAXIMUM PERCENT DEVIATION

(A=0-10, B=10-25, C=25-50, D=50-200)

	ARMY				NAVY				AIR FORCE			
CODE	A	B	C	D	A	B	C	D	A	B	C	D
TAMIL	11	3	8	1	6	8	3	0	9	1	0	0
TACIV	17	3	3	0	16	1	0	0	10	0	0	0
TPROF	11	8	4	0	12	5	0	0	6	3	1	0
TBACH	13	6	3	1	10	7	0	0	6	2	2	0
TMAST	7	11	3	2	5	11	1	0	4	5	1	0
TPHDS	6	9	5	3	6	5	5	0	3	3	4	0
TSPAC	11	5	5	2	11	4	2	0	10	0	0	0
TR+D$	7	8	8	0	6	8	3	0	1	5	3	1
TPRO$	1	4	5	8	1	3	3	10	0	1	1	2
TO+M$	3	4	3	11	0	1	6	10	2	0	0	3
TPGM$	6	11	4	2	6	9	2	0	1	5	3	1
TIH$	9	7	4	3	5	10	2	0	3	2	2	3

TABLE B

DISTRIBUTON OF DATA BASE ELEMENTS ACCORDING TO PERCENT ANNUAL CHANGE

(A=0-20,B=20-50,C=50-100,D=100-∞)

	ARMY				NAVY				AIR FORCE			
CODE	A	B	C	D	A	B	C	D	A	B	C	D
TAMIL	11	4	7	1	10	5	2	0	9	0	1	0
TACIV	18	3	2	0	16	1	0	0	10	0	0	0
TPROF	12	7	4	0	15	2	0	0	7	2	0	1
TBACH	14	4	4	1	15	2	0	0	6	2	1	1
TMAST	9	9	3	2	10	5	2	0	4	4	1	1
TPHDS	9	9	2	3	8	7	1	0	3	5	1	1
TSPAC	13	3	5	2	11	5	0	1	10	0	0	0
TR+D$	9	8	6	0	7	9	1	0	2	4	4	0
TPRO$	2	3	1	12	1	5	1	10	0	1	1	2
TO+M$	3	4	4	10	0	1	5	11	2	0	0	3
TPGM$	7	8	7	1	7	9	0	1	2	4	4	0
TIH$	10	9	3	1	6	11	0	0	3	3	1	3

FIGURE 3.15

Several other elements common to one or two of the military departments also meet the threshold defined above, i.e., TAMIL and MPROF in the Air Force, TIH$ and TPGM$ in the Army and Navy, etc.; these can be seen by inspection of the tables in Appendix C. The RDT&E Program comes close to being in the select group, but the remainder of the appropriations (Operations and Maintenance, Procurement, and Miscellaneous) are mostly in the poor category.[1] This affects the Navy laboratories most and the Air Force laboratories least, since a substantial proportion of the Navy laboratories appropriations are for other than R&D, while almost all of the Air Force program is for R&D. The amount of annual variation in patents, papers, reports, training, etc., was surprisingly moderately low in view of having only two years of data, and also in consideration of the basic unpredictability of the creative process.

The deviations from the means are considerably smaller when viewed on a departmental basis, and even more so when looked at collectively for all fifty-one laboratories. In this latter sense, only twenty-four elements have deviations more than ten percent from the mean, and half of these are less than twenty percent. The remaining twelve, i.e., those with the most variation, have deviations of less than fifty percent. These include

MILND - Military Professionals, No Degree
MILCN - Military Construction
MILPA - Military Pay and Allowances
DEPPR - Department Procurement
IHO&M - In-House Operations and Maintenance
OHO&M - Out-of-House Operations and Maintenance
IHOMA - In-House Other Miscellaneous Appropriations
OHOMA - Out-of-House Other Miscellaneous Appropriations
DEPOM - Department Funds, O&M
DEPMS - Department Funds, Miscellaneous
OTHMS - Other DoD Funds, Miscellaneous
NONMS - Non-DoD Funds, Miscellaneous

3.5 Summary Comments About the Validation of the Data Base

Much of the personnel and financial data was subject to verification by cross-checking. The validation criteria included the requirements that the various types of appropriations equal the corresponding sources of funds, that in-house and out-of-house components equal their respective totals, and that on-board personnel

[1] A considerable portion of the variations in these and other data elements is attributable to vagaries in expenditures at the laboratory level and is probably meliorated through the carrying over of funds from one fiscal year to the next.

counts did not substantially exceed authorized limits. A number of minor mistakes were found and corrected. In a few instances where it was not possible to figure out what was wrong, revised data were obtained from the laboratories in question. One exception to this was the data describing the value of scientific and engineering equipment; since these data were reported based upon different interpretations of the instruction, there remains a residual uncertainty concerning them.

Other parts of the data base which were not subject to arithmetic validation have been visually inspected a number of times. Several mistakes - most of them apparently either in recording or keypunching the data - were found in items such as patents, reports, papers, etc. In a few cases, some of these data were changed to reflect more reliable information obtained for separate purposes. There are a few occasions where the data satisfies the validation criteria, but is clearly anomalous. For example, one laboratory, with a total program in excess of fifty million dollars, reported more than half of its appropriations from miscellaneous sources (Military Pay and Allowances, In-House Other Miscellaneous Appropriations, and Out-of-House Other Miscellaneous Appropriations) for two of the three years under consideration.

The validations were carried out first for the data for fiscal year 1969, then for fiscal year 1968, and lastly for fiscal year 1967. This latter was never completely validated, in the sense that the major check sums were not satisfied. The elements having the most discrepancies are the three elements having to do with miscellaneous source of funds: Department Miscellaneous (DEPMS), Other DoD Miscellaneous (OTHMS), and Non-DoD Miscellaneous (NONMS). The principal source of the discrepancy in most cases is that Military Pay and Allowances (MILPA) was not included in the fiscal year 1967 source of funds elements.

Effect of Change

Three kinds of change, other than the normal year-to-year variations in laboratory properties, are observable in the data over the three-year period. These are: changes in the laboratories' organization, the addition of new elements, and a change in the interpretation of existing elements.

The organizational changes include closing of laboratories, consolidation with others, reassignments of portions of a laboratory's mission, and re-structuring of the laboratory's composition. The consolidation of the Naval Ordnance Laboratory (Corona) with the Naval Weapons Center (China Lake) has been cited previously. The consolidation is regarded as having been in effect from the beginning of fiscal year 1968, but the major phasing out did not begin until fiscal year 1969; consequently, the data which was obtained for fiscal year 1968 is considered as

fairly representative of NOLC for that and the preceding fiscal year.[1] In most cases, the fiscal year 1968 NOLC data have been processed separately from the NWC data; but for comparison with fiscal years 1967 and 1969, and in computing three-year averages of the laboratory properties, the Corona data have been re-combined with that of China Lake.

Two of the other laboratories undergoing substantial changes were the Ballistics Research Laboratory and the Picatinny Arsenal Laboratories. As part of a reorganization at the Aberdeen Proving Ground, the personnel ceiling of the Ballistics Research Laboratory was decreased from 1261 in fiscal year 1968 to 651 in fiscal year 1969. The total program, the R&D program, and the number of professionals were correspondingly reduced by about a third. In another instance, the personnel ceiling of the Picatinny Arsenal Laboratories decreased from 3379 in fiscal year 1967 to 2117 in fiscal year 1968, although the number of professionals remained the same. There was some reduction in appropriations, but not so extreme.

Addition of New Elements

Several elements have been added to the original data base. Eleven elements were added in fiscal year 1968; these included SEQIP, SEQNP, SEQPR, PATNT, PAPER, RPRTS, CIVGS, MILGS, CFTGS, MFTGS, and MEETS, bringing the number of elements up to sixty-nine. A seventieth element, Military Skilled Trades (MILST), was added in fiscal year 1969. The data base was further augmented in 1970 by a revision of DoD Instruction 7700.9, calling for the inclusion of four new elements:

- Research with Universities (total obligational authority for research (6.1) conducted out-of-house with colleges or universities).
- Research with Industry (total obligational authority for research (6.1) conducted out-of-house with industry).
- Exploratory Development with Universities (total obligational authority for exploratory development (6.2) conducted out-of-house with colleges or universities).
- Exploratory Development with Industry (total obligational authority for exploratory development (6.2) conducted out-of-house with industry).

[1] In terms of the principal staffing and funding elements - TPROF, TAPER, TR&D, and TPGM$ - the FY 68 data of NOLC and NWC combined is within 7% of the corresponding NWC FY 67 or FY 69 data. The FY 68 NOLC data itself ranges from 21% (for TAPER) to 38% (for TPROF) of the FY 68 NWC data.

Differences in Definition or Interpretation

A few of the elements appear to have annual variations attributable to changes in interpretation of the elements, e.g., the reporting of Scientific and Engineering Equipment. Other variables that seem to fluctuate unusually are the space elements and professionals-no-degree. Six of the Navy laboratories reported more than a fifty percent change in Laboratory Space, Administrative Space, and Other Space (although only one had a Total Space change in excess of fifty percent). Ten of the Army laboratories showed similarly large fluctuations. The Air Force laboratories were relatively stable; there was only one change in Laboratory Space greater than fifty percent, and only two in Other Space.

With the introduction of Military Skilled Trades in fiscal year 1969, a number of DoD activities - including some laboratories - showed a shift in staffing from professionals-no-degree to military skilled trades. A similar shift from professionals-no-degree to technicians probably reflects changes in staffing policies, or a reclassification of personnel.

Overall

The personnel elements score fairly high in density and stability; their major shortcoming is that they give the on-board count as of June 30 of each year. They might be more representative if they were based on an on-board count over the whole year. The facility elements such as Scientific Equipment and Laboratory Space would seem a priori to be highly associated with laboratory capability, but they are among the least reliable because of differences in interpretation and fluctuations in reporting. Variables like patents, reports, etc., can vary so much in quality and significance that their quantity is not a reliable measure of laboratory quality (although when coupled with other variables, the numbers may be meaningful). The financial elements have the most amount of variation for a number of reasons: the necessity for the sponsor to fund certain priority programs and to support others as best he can; the desire of the laboratory to carry out certain of its priority programs with such funding as can be obtained; the (sometimes large) differences between what is appropriated and what is actually spent; and the inherent ambiguity in reporting transactions of these various types.[1]

---

[1]In some of the inter-departmental transfers of funds, the type of appropriation is not always known to the receiving agency, and oftentimes is reported as miscellaneous appropriations.

Overall, while the data for any one year is subject to fluctuation from a number of different sources, over a period of years the variation will be statistically damped and major trends should become apparent. The data for fiscal year 1971 will extend the range of the base to five years, which should provide a fairly reliable basis for perceiving the directions in which properties are moving and their relative rates of change.

4. RELATIONSHIPS BETWEEN RATINGS AND LABORATORY PROPERTIES

4.1 Background

Scope and Purpose

Within the Department of Defense, many different types of appraisals are regularly made - supervisory evaluations, program evaluations, special appraisals, committee visits, and the natural competition of laboratories for important programs. Most of these techniques are subjective in nature and lack a quantitative basis, particularly for comparisons among laboratories with widely differing missions and technical orientation. To rectify some of the deficiencies in these appraisal systems, the laboratory resources data base was developed to provide comparative statistical and trend data on the characteristics and performance of laboratories. It was felt that the utility and significance of these data might be improved if they could be related to the comparative technical competence or quality of laboratories. The peer ratings described in Chapter 2 were obtained for this purpose.

The present report is limited to studies made to examine elementary relationships between the data base and the peer ratings. These represent the first phase of a larger study whose "ultimate goal is to give managers of DoD laboratories a greater insight into research management and organization, and to help them in efficiently using data on laboratory properties, performance and their relationships, for purposes of self-evaluation and self-improvement." [2] Factors which might be considered in subsequent studies would include geographical location, management policies, organizational structure, professional attitudes, leadership patterns, and so on. An analysis of these additional variables would provide more thorough understanding of the management actions and policies that influence laboratory productivity. The present study does not address relationships so complex as these, but rather is limited to an examination of the institutional characteristics described in Chapter 3.

Methodology

The present work relies primarily upon the use of simple linear correlation to measure the degree of association between the ratings and the data. Rank-order methods and multiple regression have also been used - the former for intuitive and corroborative purposes, and the latter to examine the relationship between the peer ratings and groups of data elements.

The rank-order procedures consist of inspecting the properties of the laboratories according to the magnitude of some characteristic, e.g., peer rating or any of the elements of the data base. In some instances the coefficient of rank-order correlation has been used to indicate the degree of association; in others it has been convenient to illustrate the importance of a particular property by displaying its proportion among high-rated laboratories versus the lower-rated laboratories.

I have used the term "correlation" to mean a measure of the association between two variables, and the term "regression" to signify a relationship wherein one variable is a linear function of one or more other variables. I have been more interested in the coefficient of multiple correlation resulting from regression than in the coefficients of the regression equation; and more in the relative change of the multiple correlation coefficient than in its absolute value.

Rationale

For the most part, no attempt has been made to establish a causal relationship between the peer ratings and the quantitative properties of laboratories, partly because of the indeterminancy of the time element and partly because it is not certain what is cause and what is effect. Is a laboratory highly rated because it obtains a greater than average amount of research money? Or does it receive the money because it was already highly esteemed? In either event, the relationship is not likely to be a simple one between money and rank, but rather depends upon what the laboratory has done with the money (or other laboratory resources).

The other uncertainty - the time element - is coincident with the above. Which year of the data base should be used for comparison? Or should one strike an average? Did a participant know a laboratory as it was in 1969 - or was he basing his opinion on knowledge obtained in 1968, or 1967, or earlier? Roman points out that "impressions of an organization tend to persist even though unsupported or gleaned from a distant past. Earlier weaknesses may have been rectified or current performance may not measure up to former accomplishment, but reputation labels stuck on organizations are hard to dislodge." [6][1]

Questions of this sort are neither exhaustively enumerated nor satisfactorily answered; but rather have been looked at in a variety of ways: from over, and under, and sideways. The principal results have been displayed in the various

---

[1]From **RESEARCH AND DEVELOPMENT MANAGEMENT: The Economics and Administration of Technology.** Daniel D. Roman. Copyright (c) 1968 by Meredith Corporation. By permission of Appleton-Century-Crofts, Educational Division, Meredith Corporation.

appendices attached, with the expectation that the reader will interpret them in the light of this own experience.

4.2 Development of the Study

The study of relationships between the peer ratings and the resource data base was undertaken shortly after the peer ratings were obtained, and was initially conducted by Steve Smith while assigned to the Office of Laboratory Management. Statistical analyses and computational support during this phase of the work were provided by Bert Levy and John Marsh at the Army's Harry Diamond Laboratories.

At the time of these studies, only the data for fiscal years 1967 and 1968 were available. The analyses were conducted using the FY 68 data, which at that time was available for only fifty-two of the fifty-five non-medical laboratories.[1] In addition to examining the relationship between the peer ratings and the elements of the data base, the investigators also wished to explore associations between the ratings and various normalized elements, to determine whether proportional variables might be more highly correlated with the ratings than the standard variables. For example, which is more significant - the amount of total program, or the proportions of the amount used for RDT&E, O&M, etc.? Is the ratio of PhD's to professionals more related to laboratory reputation than either of the elements individually? To look into questions such as these, thirteen different factors were used to normalize the data. Correlations between the peer ratings of the fifty-two non-medical laboratories and each of the thirteen normalized data sets were computed and compared with the correlation between the peer ratings and the unnormalized data; the results are shown in Appendix D. (This does not show the complete set of variables, but is a selected subset of those properties considered by the author to be most germane.)

In all but three of the cases shown, the unnormalized data exhibits a higher absolute correlation with the peer rating than do any of the normalized sets. The exceptions are In-House Operations and Maintenance (IHO&M), Professional Civilians

[1]The data for the Naval Ordnance Laboratory (Corona) had been combined with that of the Naval Weapons Center. The data for the Mine Defense Laboratory and the Marine Engineering Laboratory had been combined with that of the Naval Ships Research and Development Center. These data were later obtained separately and used in subsequent studies.

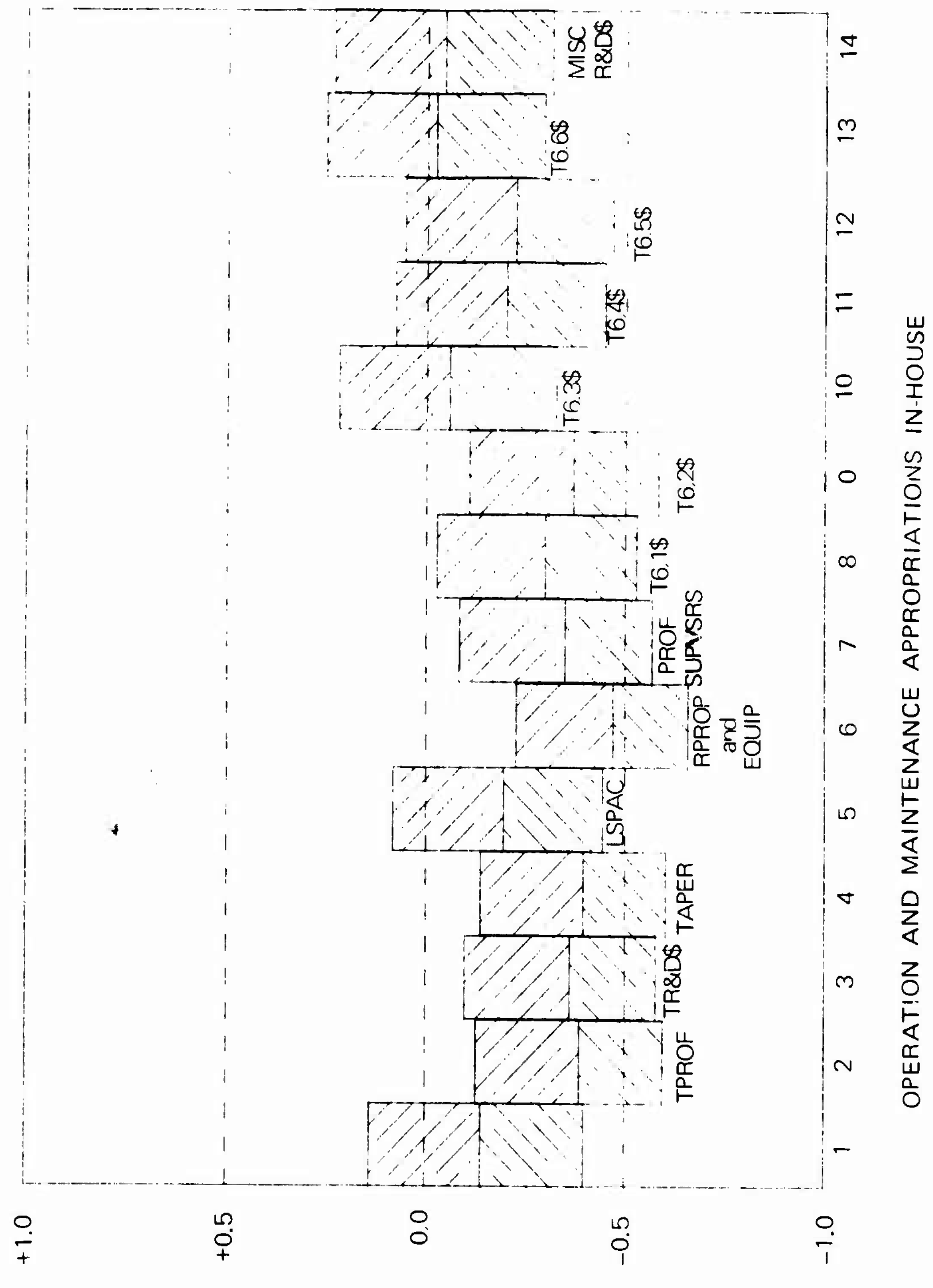
With No College Degree (CIVND), and Civilian Graduate Students (CIVGS). In each case, the exceptional values are negatively correlated with peer ratings. In one case, Technical Reports, the unnormalized correlation is approximately equal to the correlation of the data normalized by Total Research, but opposite in sign. In another case, Out-of-House Research, there is little variation between any of the correlations, normal or unnormalized. The correlations for IHO&M are shown in Figure 4.1; the diagram also shows the 95% confidence interval bounding the correlation. Case 1 is the unnormalized data; the various normalization factors are described in Appendix D.

The correlation of the unnormalized variables is most generally positive; and in several cases it is considerably larger than any of the normalized variables, viz, In-House Research, Civilians with Master's Degree, Civilians with Doctor's Degree, and Papers Published. The correlations between these variables and the peer ratings range from .5 to .6; the correlations between the variables range from .64 to .85. As noted under the discussion of the correlations between variables in Section 3.3, these particular variables (CIVPH, IH6.1, and PAPER) are highly inter-related.

Development of the Present Work

Shortly after the above analyses had been completed, the fiscal year 1969 data became available. Dr. Smith experimented with a number of combinations of the elements of the fiscal year 1969 data base in a search for a normalization factor that might better account for variations in the peer ratings. Programming assistance and computational support were obtained from Rich Hein and Frank Reynolds in the Air Force Information Systems Division in the Pentagon. Correlations between these normalized data and the peer ratings of the seventy-seven DoD laboratories for which data was available yielded results that were less significant than correlations using the unnormalized data. The author carried these experiments a step further by utilizing a "ratio" data base that had been created two years previously by Locher and Haberman [7] using the fiscal year 1967 laboratory properties data obtained by Anderson. The ratio data base consisted of thirty-nine of the regular elements normalized by variables such as total professionals, total research dollars, total RDT&E dollars, etc.

The ratio data base combinations were applied to the fiscal year 1969 data for all the laboratories for which data were available; again, the results were not as highly correlated with the peer ratings as were the unnormalized variables. It was



CORRELATION WITH RANKING
AND 95 PERCENT CONFIDENCE LIMITS

FIGURE 4.1

decided to divide the laboratories into two groups: the medical and the non-medical (this had also been done previously with the fiscal year 1968 data). The principal correlations are shown in the three left-hand columns of Figure 4.2. Those that have been selected for illustration consist of the correlations that are most significant (at the .01 level), together with others of special interest. The numbers in parentheses indicate the number of laboratories having non-zero values of the particular property.[1] Where there were less than five laboratories having a particular property, the correlation is not shown.

The correlation was computed using only r,x pairs where the x-value is different from zero. Only thirty-one of the medical laboratory elements were more than 50% dense (i.e., were possessed in some degree by more than one-half of the laboratories) compared with fifty-nine for the non-medical laboratories. The staffing elements (TAMIL, TACIV, etc.) of the medical laboratories show a considerably higher correlation with the ratings than do those of the non-medical laboratories, but they are about at the same level of significance. In most cases, the correlation of the medical laboratories combined with the non-medical laboratories tended to be lower than either of the two separately. One reason for this is the considerable difference in magnitudes of the medical laboratory properties versus those of the non-medical laboratories. Except for a few instances, mainly involving military personnel or those with advanced degrees, the characteristics of the non-medical laboratories are many times the size of the medical laboratories, but their ratings span almost the same range.

Another subdivision of the laboratories was made according to military departments. The principal correlations for the same elements that were shown previously are listed in the three right-hand columns of Figure 4.2. It was apparent that there were significant differences among the three services; consequently it was further decided to subdivide the medical and non-medical laboratories according to military departments. This gave rise to eight combinations of laboratories: (Medical/Non-medical; Army, Navy, Air Force, DoD). Correlations were computed for each combination using both the regular and the ratio variables. The main consequences of this experiment were (1) a decision to separate the medical from the non-medical laboratories, and in the present study to concentrate on the latter; (2) to process the laboratories of the three military departments in three separate groups.

[1] In a few instances the numbers of medical and non-medical laboratories add up to one less than the DoD total. This is because a portion of the data for one of the Navy laboratories was missing from the data base at the time the correlations were computed.

PRINCIPAL CØRRELATIØNS (FISCAL YEAR 1969)						
	ARMY, NAVY, + AIR FØRCE LABØRATØRIES			MEDICAL AND NØN-MEDICAL LABØRATØRIES		
	MEDS	NØN-MEDS	DØD TØTAL	ARMY	NAVY	AIR FØRCE
TAMIL	.56(23)	.33(54)	.41(77)	.46(35)	.34(26)	.41(16)
TACIV	.64(23)	.43(54)	.30(77)	.17(35)	.43(26)	.40(16)
CIVBS	.58(21)	.39(54)	.23(75)	.01(34)	.43(26)	.24(15)
CIVMS	.68(19)	.55(54)	.38(73)	.18(33)	.57(25)	.41(15)
CIVPH	.72(20)	.57(52)	.48(72)	.30(32)	.62(25)	.43(15)
MILPH	.59(22)	-.01(40)	.30(62)	.44(27)	.27(19)	.36(16)
CLASS	.63(23)	.43(54)	.30(77)	.17(35)	.44(26)	.36(16)
TECHS	.61(23)	.40(54)	.33(77)	.24(35)	.47(26)	.33(16)
ØWNED	.26(7)	.37(42)	.32(50)	.53(17)	.37(18)	.17(14)
LEASD	.47(13)	.42(22)	.37(35)	.63(16)	.36(14)	.24(5)
RPRØP	.54(21)	.46(51)	.36(73)	.12(32)	.48(24)	.43(16)
LSPAC	.60(23)	.45(53)	.35(77)	.12(35)	.47(25)	.37(16)
EQUIP	.38(23)	.49(52)	.37(76)	.18(34)	.54(25)	.29(16)
SEQIP	.36(23)	.38(51)	.30(75)	.20(34)	.40(26)	.23(16)
SEQPR	.56(21)	.53(45)	.41(67)	.32(31)	.58(21)	.08(14)
MILPA	.52(23)	.38(34)	.45(58)	.57(24)	.31(23)	.61(10)
HØUSE	.48(22)	.55(49)	.37(72)	.17(32)	.51(25)	.38(14)
IHR+D	.64(23)	.50(54)	.38(77)	.16(35)	.61(26)	.48(16)
IH6.1	.60(17)	.57(48)	.49(65)	.52(32)	.59(24)	.55(9)
IH6.2	.47(23)	.52(51)	.38(74)	.21(34)	.56(26)	.14(14)
IHØ+M	.--(4)	-.32(38)	-.31(43)	-.25(22)	-.34(16)	-.24(5)
DEPRD	.60(23)	.43(54)	.30(77)	.08(35)	.54(26)	.40(16)
DEPMS	.55(22)	-.07(33)	-.01(55)	.22(24)	.30(23)	.67(8)
NØNRD	-.36(6)	.52(23)	.46(29)	.48(14)	.54(8)	.41(7)
NØNMS	. (3)	.64(18)	.62(21)	.25(8)	.86(12)	.--(1)
PATNT	.23(8)	.46(46)	.36(54)	.19(24)	.60(18)	.08(12)
PAPER	.52(21)	.57(54)	.52(75)	.46(34)	.66(25)	.49(16)
RPRTS	.55(22)	.24(54)	.18(76)	.02(34)	.28(26)	.21(16)
CFTGS	.32(6)	.31(45)	.26(51)	.22(21)	.67(17)	.27(13)
MEETS	.54(23)	.49(53)	.43(76)	.26(35)	.47(26)	.65(15)

FIGURE 4.2

The decision to examine the non-medical laboratories first was motivated principally by a parallel effort initiated by the Office of Laboratory Management; this was the REFLEX (for **RE**sources **FLEX**ibility) study. Characteristically, the laboratories have been subject to both manpower and fiscal constraints in the management of their technical programs. The REFLEX project is a two- to three-year demonstration study being conducted at ten in-house physical sciences and engineering laboratories to determine whether financial controls alone can be used in place of the combined fiscal and manpower ceiling controls now employed. The key to evaluating this project is in determining how the removal of manpower ceilings affects the performance of the laboratories involved.

The results obtained up to this time were summarized in tabular form for inclusion in a paper being written by Glass [3]. Shortly thereafter, I enlisted the support of the statistical and programming staffs of the Naval Weapons Laboratory for the utilization of existing statistical programs and for the development of additional computer programs to aid in processing the data. Statistical consultation and support were provided by Marlin Thomas and Gary Gemmill; most of the programming was done by or under the supervision of Ray Brancolini.

Presentation of Results

The results obtained from the use of these various programs are described in Chapters 5, 6, 7, and 8. Chapters 5 and 6 describe various analyses conducted to examine the sensitivity of the correlations to variations in the data; the discussion became so lengthy that it has been separated into two chapters. The first of these covers the basic correlations between the peer ratings and the laboratory properties, using the data for fiscal year 1968; examines the correlations for other fiscal years of the data base, and for the average of all three fiscal years; and considers the variations in correlations according to different rater groups. Chapter 6 looks at the dependency of the correlations upon extreme points, and upon the largest and the highest-rated laboratories in each military department; examines the results of logarithmically transforming the data; and considers the correlations obtained by using various ratio variables. Chapter 7 looks at the relationships between the rank-ordering of the peer ratings and various normalized and unnormalized combinations of the laboratory properties, and also focuses upon relationships between the properties of high-rated versus low-rated laboratories. Chapter 8 utilizes regression analyses to examine multivariate associations between the peer ratings and the elements of the laboratory properties data base.

5. CORRELATION ANALYSIS (I)

5.1 Correlation Analyses

Having distinguished between the medical and non-medical laboratories, and having elected to focus on the latter, there were still a number of questions to be considered with respect to the analytical conduct of the study. Should all the non-medical laboratories be used, or only a part? How does the association vary according to rater groups? What variation is there between the peer ratings and the different years of the data base? How does the distribution of a variable among the laboratories affect the correlation? One way to examine these and similar questions would be to compute the correlation coefficient for these various conditions, in order to obtain both a qualitative and quantitative measure of the variation.[1]

To carry out these various examinations, I used a computer program prepared by David Wolper at the Naval Weapons Laboratory. This program computed the correlation for each military department as well as for DoD as a whole. It permitted the computation of correlation coefficients with and without zeros, as well as providing a measure of the distribution of the laboratory variables. With minor modification, it was used to compute the correlation between pairs of variables (i.e., for cluster analyses), the correlation between a variable and the peer ratings, and the correlation between the product, sum, or ratio of a pair of variables and the peer ratings. This latter feature was used to examine combinations of products and ratios using the average of the fiscal year 1967, fiscal year 1968, and fiscal year 1969 data.

I considered the capability to view the distribution of laboratory values an essential feature of the correlation analyses. In generating the combinations of sums, products, and ratios, it would be necessary to know if a high correlation was unduly influenced by an extreme point, as well as whether the variables (particularly ratios) were statistically separable. The program previously used at the Pentagon had provided a one-page graph of the ratings versus each laboratory property, as is illustrated in Figure 5.1.[2] These provided a snapshot view of the association - an

---

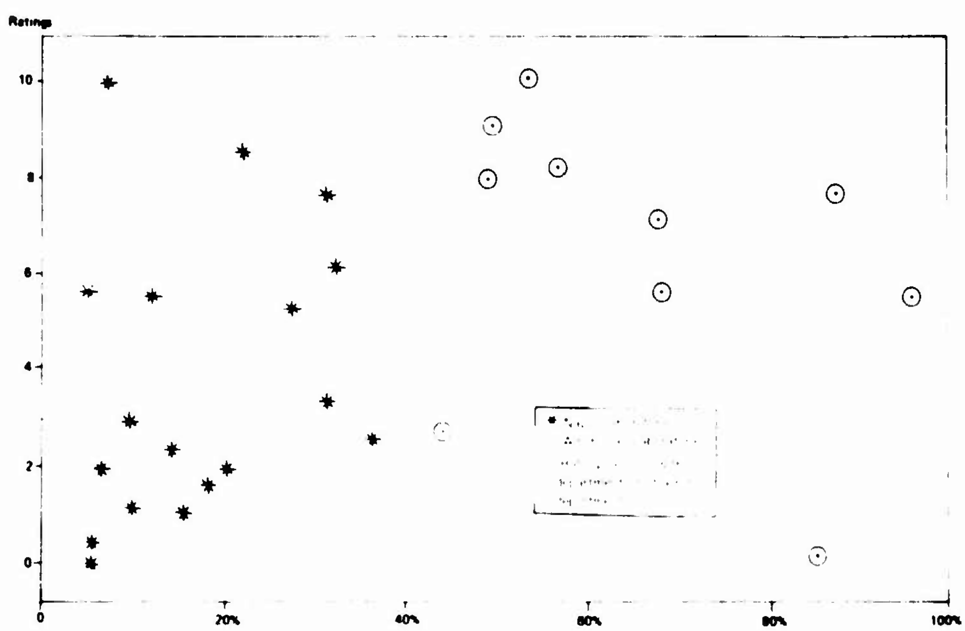
[1] In trying to quantitize relationships between the peer ratings and the data base of laboratory properties, it is possible to view the ratings as being dependent on the properties (or vice versa), or simply to view them as two sets of variables without specifying their inter-dependency. In a subsequent section, where many properties are being simultaneously associated with the laboratory ratings, it is assumed that the rating is dependent upon the properties. In the present section the ratings are associated with the laboratory properties one at a time, and while there is an implicit dependency of ratings upon properties, the mathematical process employed is the technique of simple correlation.

[2] See discussion on page 3-10 and 9-3. The figure shows that none of the Navy laboratories exceeds 40% Out-of-House R&D; none of the Air Force laboratories is less.

obvious pattern would be quickly discernible. However, with many thousands of such associations to examine, it would be desirable to have a more compressed measure of the relationship. I considered using measures such as the standard deviation and relative dispersion, but neither of these seemed satisfactory.

In lieu of developing a satisfactory measure of these conditions, I elected to partition the range of the laboratory variables into ten equal zones and to visually inspect the distribution (although this would not provide for recognition of the case where a significant correlation is masked by an extreme point). A listing of the output of the correlation program for fiscal year 1969 is shown in Figures 5.2 and 5.3; this corresponds to the general format used in Appendices A through K.

Figure 5.2 lists the correlations for Army, Navy, Air Force, and DoD under the sub-headings labeled R (for "rho"). In the example shown, zeros are included in the computation of the correlation coefficient, but the number of laboratories having non-zero values of the elements is shown under the column headed "N", except for the DoD column, which shows the total number of laboratories under consideration.



Proportion of OHR&D to TPGMS, FY 68
FIGURE 5.1

1968 DATA		ALL RATERS		REGULAR				COUNT ZEROS	
INDEX	CODE	ARMY		NAVY		AIR FORCE		OOD	
		R	N	R	N	R	N	R	N
75	MPROF	0.176	23	0.133	18	-.143	10	0.141	51
76	CPROF	0.254	23	0.822	18	0.051	10	0.421	51
77	TPROF	0.254	23	0.801	18	-.043	10	0.436	51
78	TBACH	0.218	23	0.683	18	-.452	10	0.346	51
79	TMAST	0.362	23	0.881	18	0.416	10	0.577	51
80	TPHDS	0.368	23	0.799	17	0.621	10	0.578	51
81	TAPER	0.254	23	0.689	18	-.074	10	0.399	51
82	TSPAC	0.222	23	0.651	18	0.094	10	0.448	51
83	TR+D$	0.171	23	0.779	18	-.092	10	0.365	51
84	TPRO$	0.215	18	-.222	18	-.209	3	-.031	51
85	TO+M$	0.040	19	-.319	18	-.372	5	-.166	51
86	TPGM$	0.238	23	0.488	18	-.157	10	0.315	51
87	TIH$	0.222	23	0.774	18	0.213	10	0.439	51
88	IH1-2	0.416	22	0.707	18	0.174	10	0.563	51
89	IH1-3	0.334	23	0.794	18	0.168	10	0.546	51
90	IH1-4	0.289	23	0.808	18	0.157	10	0.510	51
91	OH1-2	0.155	20	0.269	18	0.102	10	0.234	51
92	OH1-3	0.044	21	0.302	18	0.064	10	0.179	51
93	OH1-4	0.030	21	0.349	18	-.177	10	0.140	51
94	TDEP$	0.190	23	0.440	18	-.140	10	0.264	51
95	TOOOD	0.389	21	0.361	16	-.197	8	0.325	51
96	TNDOD	0.410	16	0.588	13	0.387	6	0.443	51
97	T6.1$	0.464	21	0.713	18	0.484	9	0.526	51
98	T6.2$	0.256	22	0.700	18	-.101	9	0.374	51
99	T6.3$	0.034	17	0.561	17	-.040	8	0.131	51
100	T6.4$	0.032	17	0.400	16	-.827	6	0.073	51
101	T6.5$	0.055	19	0.182	18	0.171	6	0.048	51
102	T6.6$	0.078	14	0.618	13	-.146	3	0.218	51
103	T61-2	0.336	22	0.729	18	0.163	10	0.513	51
104	T61-3	0.166	23	0.793	18	0.115	10	0.402	51
105	T61-4	0.131	23	0.739	18	-.128	10	0.343	51
106	TOH$	0.226	22	-.066	18	-.242	10	0.127	51
107	ACRES	0.401	20	0.408	18	0.175	10	0.297	51
108	SEQAS	0.215	22	0.792	18	0.416	9	0.541	51
109	TOTNO	0.002	21	0.150	18	-.102	9	-.010	51

FIGURE 5.2

1968 DATA   ALL RATERS   REGULAR   CØUNT ZERØS

ARMY DENSITY	M/M	NAVY DENSITY	M/M	AIR FØRCE DENSITY	M/M	INDEX
843321--11	1E 2	7532-----1	3E 1	612------1	9E 0	75
936211---1	6E 1	3225-2-121	8E 0	32-111---2	5E 0	76
935311---1	5E 1	232412--13	7E 0	21211--1-2	3E 0	77
A242211--1	5E 1	22242111-3	1E 1	1-3212---1	1E 1	78
95341----1	1E 2	243133-1-1	1E 1	332----1-1	5E 0	79
94321--121	7E 1	93311----1	2E 2	51-21----1	3E 1	80
83452----1	8E 1	33422111-1	1E 1	32-21--1-1	4E 0	81
93213112-1	1E 2	97--1----1	3E 1	1121--31-1	1E 1	82
A7-11----1	1E 2	5342--1-21	1E 1	1-5-11-1-1	6E 0	83
A--1----11	6E 4	9411---1-2	6E 2	8-----1--1	2E 4	84
A34--111-1	7E 3	543111-111	6E 1	81-------1	6E 3	85
A5113----1	1E 2	336-122--1	2E 1	12221---11	7E 0	86
A362-1---1	1E 2	437-1-1--2	1E 1	112121---2	1E 1	87
A223121--1	2E 4	A411-----1	1E 1	22211----2	9E 0	88
A324-----1	8E 1	83321----1	1E 1	22211----2	9E 0	89
A4321----1	1E 2	833-111--1	2E 1	22211----2	9E 0	90
A71-2----1	2E 4	62311111-2	3E 1	213-1-1-11	2E 1	91
A4-------1	7E 4	75-11-2-11	9E 1	322-11---1	1E 1	92
A6-------1	1E 5	A112-1-1-1	4E 1	223-11---1	1E 1	93
A52-3----1	1E 2	31531211-1	1E 1	2-3-3----2	6E 0	94
A211-----1	3E 4	A21-1---11	3E 4	6-11---1-1	1E 4	95
A2121----2	1E 3	A1-------1	1E 4	5-13-----1	7E 2	96
A43-1-1--1	1E 4	A41------1	2E 2	5112-----1	3E 4	97
A433-11--1	3E 4	8413--1--1	9E 0	12112---21	5E 4	98
A2-------1	7E 4	72221-2-11	2E 4	4131-----1	2E 4	99
A71--1---1	4E 4	A3----11-1	2E 4	72-------1	2E 4	100
A33------1	1E 4	531123-2-1	2E 1	8--1-----1	6E 3	101
A131-----1	3E 4	53251-1--1	1E 4	8-----1--1	4E 3	102
A4413----1	4E 4	A23-1----1	1E 1	12-3---112	8E 0	103
A81------1	1E 2	8222111--1	1E 1	3-3-2-1--1	6E 0	104
A72------1	2E 2	8222--1-21	1E 1	2-4-2--1-1	6E 0	105
A22-1-2--1	1E 5	8123-11-11	4E 1	12212----2	1E 1	106
A321-----1	1E 4	A--------1	6E 4	9--------1	6E 3	107
A441--1--1	5E 3	94-21---11	6E 1	221-11-1-2	2E 3	108
A4321----1	1E 2	523322---1	3E 1	24-1-2---1	4E 1	109

FIGURE 5.3

An option exists to compute the correlations omitting the laboratories that have zero values of an element; in this case, if there are only three or less than four non-zero variables, the correlation is not computed for that element.

The relative densities are shown on the right hand side of the output, as shown in Figure 5.3.[1] These are shown for each of the three services, but not for DoD (an earlier version listed DoD but not the services; this is shown in Tables 6 and 7 of Appendix C). The density distribution is divided into ten zones, which represent the difference between the largest and the smallest values of the particular element. A letter "A" in one of the zones means that ten or more laboratories are in that zone; otherwise the number shown is the actual number of laboratories in that zone. A dash means no values in that zone.

The range of the elements is given by the four-character expression following the densities for each service. If there are no zeros in the data (which must be determined by reference to the left-hand page) the value shown is the ratio of the largest element to the smallest. If there are zeros, the expression is the magnitude of the largest. The expression 3E02 is read as 3×10^2; negative exponents are shown as complements of 100 (anything below 1×10^{-7} is recorded as F93).

As an illustration of the interpretation of the tables shown in Figure 5.2 and 5.3, consider element number 85: T0+M$. This is the sum of In-House and Out-of-House Operations and Maintenance. The correlation between this and the twenty-three Army laboratories (four of which had none) is .040. The value of the largest was on the order of $7,000,000 (the units shown are in kilobucks). The correlation between the peer ratings and T0+M$ for the eighteen Navy laboratories is -.319; the ratio of the largest to the smallest is about 60 (the distribution is relatively uniform, not as lopsided as some of the others). Only half of the ten Air Force laboratories had Operations and Maintenance Appropriations in fiscal year 1968; the correlation of -.373 includes the five who had none. The largest value was 6E 3, or approximately 6000 kilobucks. Since the lowest zone includes nine laboratories, five of which were zero, this implies that three of the laboratories had less than approximately 600 kilobucks each. (The correlation using only the five that actually had O&M dollars is .872; the distinction between computing the correlations with and without zeros will be considered below.)

---

[1] In the appendices, the pages have been presented face to face, instead of back to back as in Figures 5.2 and 5.3. Hence, when Figure 5.2 is referred to as the "left-hand page", and Figure 5.3 is designated as the "right hand page", the orientation is with respect to that shown in the appendices.

5.2 Statistical Significance of Correlations

It was stated in discussing the correlations in Figure 4.3 that they "were significant at the .01 level". That is another way of saying: suppose the correlation were really zero, then the probability of obtaining a correlation coefficient as large as the one observed is less than .01. This isn't a great deal of help, since it doesn't tell us what the probability is of having obtained one as large as we have; but it does provide a relative measure for comparing different correlation coefficients.

The test for significance depends upon two parameters: the level of significance desired, and the number of points in the sample. The following criteria may be helpful in determining the approximate significance of the various correlations presented in this report. If rho (R) is to be significant at the .05 level (95% confidence), then the number of points (N) must be greater than or equal to the number shown below.

R	N
.9	5
.8	6
.7	8
.6	11
.5	16
.4	24

This says right off that to be significant at the .05 level, the Air Force correlations must be greater than .6, the Navy's greater than .5, and the Army's greater than .4 - since there are at most 10 Air Force laboratories, 17 Navy, and 23 Army. (This is a necessary but not a sufficient condition for significance at the .05 level.) The values of R, N, shown above were derived from Table VII in reference [8] parts of which are reproduced in Figure 5.4[1] (adjusted for two degrees of freedom).

It is also possible to compute confidence intervals based on various levels of significance. In fact, the initial version of the correlation program computed the upper and lower bounds of the 95% confidence interval containing the correlation; samples of the output of this program are included in Appendix I. The program was used for some of the early correlation analyses, but was abandoned in favor of the program described in the preceding section.

---

[1] From **Statistical Tables for Biological, Agricultural, and Medical Research**, by R. A. Fisher and F. Yates. Copyright (c) 1963 by R. A. Fisher and F. Yates. By permission of Oliver and Boyd, Ltd., Edinburgh.

N	Significance Level .10	.05	.01
4	.9000	.9500	.9900
6	.7293	.8114	.9172
8	.6215	.7067	.8343
10	.5494	.6319	.7646
12	.4973	.5760	.7079
14	.4575	.5324	.6614
16	.4259	.4973	.6226
18	.4000	.4683	.5897
20	.3783	.4438	.5614
22	.3598	.4227	.5368
32	.2960	.3494	.4487
42	.2573	.3044	.3932
52	.2306	.2732	.3541
62	.2108	.2500	.3248
72	.1954	.2319	.3017

FIGURE 5.4
Values of the Correlation Coefficient For Different Levels of Significance

5.3 Zeros vs. Non-Zeros

At the beginning of the study, using the computer programs developed by the Air Force group at the Pentagon, the correlations were being computed using only non-zero values of the laboratory properties. For example, if ten of fifty laboratories had none of a particular element, the correlation between that element and the peer ratings was computed using only the data from the forty laboratories with non-zero values.

Initially it seemed to me that the correlations computed in this manner might be a misleading measure of the actual degree of association between the peer ratings and the laboratory properties, especially when looking at the relationships on a service basis - (the effect of ignoring two zeros in ten could be more misleading than that of ignoring ten zeros in fifty). For example, the correlation between the peer ratings of Navy laboratories and the amount of miscellaneous funding they received from non-DoD sources in 1969 is .912 for the twelve laboratories that actually recived funds, but drops to .584 when all seventeen laboratories are considered. The two sets of data are shown in Figure 5.5; the non-zero values are indicated by the solid sets; the zero values are the hollow ones on the vertical axis.

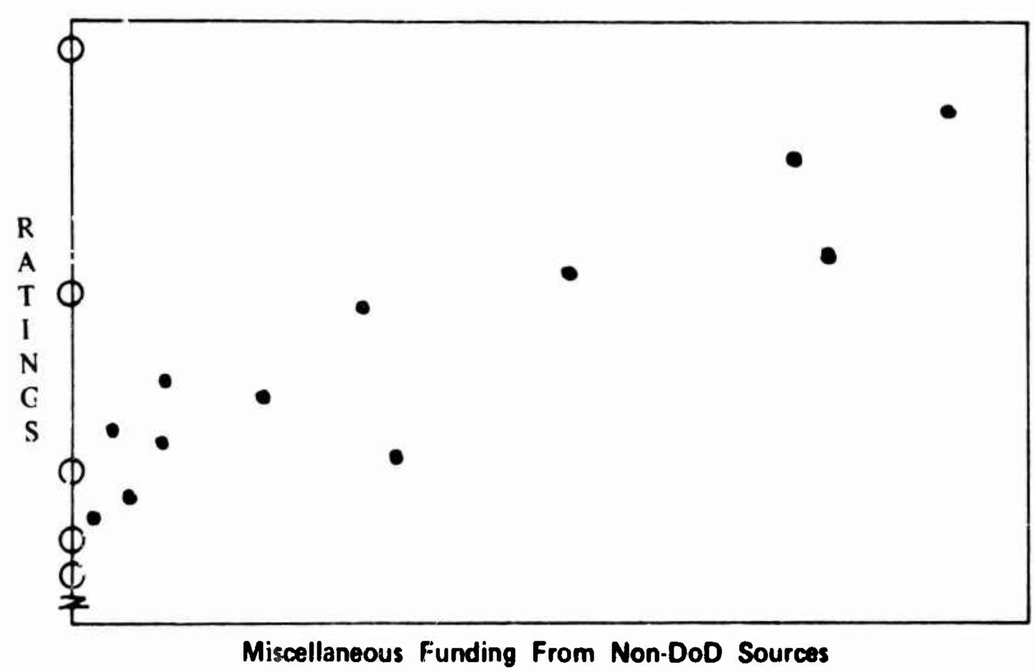


FIGURE 5.5

However, after looking over many graphs of the type illustrated in Figure 5.5, it seemed that two entirely different sets of information were pictured, and that zeros were extraneous to the situation. This feeling was further reinforced by looking at tabulations of the basic data, such as that for total research dollars for operational systems support in the Navy laboratories for fiscal year 1968 shown in Column 1 of Figure 5.6. Five of the laboratories had no dollars in this category; the abrupt transition from zero dollars to an amount in excess of two million seemed to justify computing the correlation principally for the non-zero values. With this rationale, I adopted the view that the omission of zeros was more appropriate than their inclusion, and that what should be examined was the relationship between the peer ratings and the laboratory properties for those laboratories that had non-zero values of the properties. Further, combining the zeros with the non-zeros might mask whatever significance was implied by their presence or absence.

In Dollars	In Millions	Percent of Total
18,897,000	19	21
12,533,000	13	14
7,657,000	8	9
7,034,000	7	8
6,922,000	7	8
6,897,000	7	8
6,487,000	6	7
6,469,000	6	7
5,000,000	5	5
3,880,000	4	4
3,137,000	3	4
2,394,000	2	3
2,184,000	2	2
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0

FIGURE 5.6
Total 6.6 Dollars in
Navy Laboratories FY 1968

However, the question of whether or not to include zeros arose again in the discussion of how to measure the association of the peer ratings with several properties simultaneously. In the computation of the coefficients of the multiple regression equation, it would be necessary, because of the mathematical procedure, to consider all values of a laboratory property, zeros as well as non-zeros. It was therefore decided to compute the correlations both ways - with and without zeros. Additionally, this would provide a measure of the amount of imbalance that might be present, i.e., the extent to which the two sets (zeros and non-zeros) deviate from a common trend line.

Part of the zeros/non-zeros uncertainty results from whether or not the zeros are numerical zeros or logical zeros; that is, are the zeros just the lower bound of a set of numerical values (0,1,2,···), or are they separators in a have/have not situation such as was shown in the left-hand column of Figure 5.6? The column of figures seems to indicate a clear discontinuity between the zero and the non-zero values. However, when we look at these same data in units of millions, rounded to the nearest million, what had appeared to be a distinct separation is now a barely discernible incremental step from zero in the relative scale of millions of dollars; the same phenomena can be more generally expressed in terms of percents. Much of the other financial data show an even smoother transition from zero when the values are expressed as percents of the total. Other than the numerical considerations (i.e., in the regression analysis), it would seem that the most reasonable for the inclusion or exclusion of zeros would be whether a laboratory has a choice in the matter: if a laboratory is prevented (by service policy or for other reasons), from having a particular property, then it might be more appropriate to omit it in correlating the ratings with the properties.

The correlations shown in Appendix E did not take this matter of laboratory choice into account. Instead, they were computed both ways - with zeros (COUNT ZEROS) and without zeros (SKIP ZEROS), using the data for fiscal year 1968. Some of the principal differences between the two sets are shown in Figure 5.7. The major differences occur in those instances where the variables are most sparse, as is to be expected. In almost all cases, the correlations including zeros are less than those computed excluding zeros; an exception shown in Figure 5.7 is non-DoD miscellaneous funds for the Navy; when zeros were counted, the correlation increased considerably.

Figure 5.8 lists the elements for which there is a substantial difference in the correlations of the fiscal year 1968 data with and without zeros. Only the elements having a correlation of .2 or greater are included; this includes the properties shown

Service	Element	Skip Zeros	Count Zeros	Number of Values
Army	LEASD	.862	.535	5
	OTHPR	.649	.351	5
	OTHOM	.911	.098	4
	MILGS	.551	.200	12
Navy	MILCN	.874	.637	5
	NONRD	.864	.533	6
	T6.6$	.438	.618	13
	NONMS	.526	.330	11
Air Force	MILPA	-.764	.114	4
	NONRD	.872	.495	4
	DEPOM	-.811	-.372	5
	TNDOD	.702	.387	6

FIGURE 5.7
Zero/Non-Zero Differences in FY 68 Correlations

	Army	Navy	Air Force		Army	Navy	Air Force
MILND	X	X	X	OHMPE			X
MILPH		X		NONRD	X	X	X
WGBRD			X	OTHPR	X		
OWNED	X			DEPOM			X
LEASD	X		X	OTHOM	X		
RPROP	X			DEPMS	X		X
EQUIP	X			OTHMS	X	X	
SEQNP	X		X	NONMS	X	X	X
IHO&M			X	MILGS	X	X	
MILCN	X	X		CFTGS	X	X	X
IHOMA	X			MFTGS			X
HOUSE	X		X	TO&M$			X
IH6.3			X	TNDOD			X
OH6.6	X			T6.6$		X	
MILPA	X		X				

FIGURE 5.8
FY 68 Correlations Where Ratio of Rho-Squares > 2

in Figure 5.7. The definition of "substantial" is taken to be a ratio of two or more between the squares of the correlations.

The fiscal year 1968 properties having the largest correlations with the peer ratings are shown in Figure 5.9. These are taken from the SKIP ZEROS tables of Appendix E, and consist only of values with a correlation coefficient greater than .400, except for the Navy, where the threshold value is .775. (This latter value was chosen in order to focus on the more significant correlations; 54 of the Navy elements exceeded the threshold of .400.) Where two of the services have a principal element in common, the correlations are shown for all three services; this occurs in the first twelve lines of the figure. Where the Navy values exceed those for either of the other two services being shown, the Navy values are also shown; these are lines 5-12.

5.4 Variation by Fiscal Year

Initially the correlations were computed using the data for fiscal year 1969. This was done partly because the fiscal year 1968 data had already been used (in the analyses at the Harry Diamond Laboratories), partly because the fiscal year 1969 data was the first to be re-validated, and partly because the fiscal year 1969 data was initially more complete than the fiscal year 1968 or fiscal year 1967 data. The only omission in the fiscal year 1969 data was the data for NOL Corona. The Corona data was included with the NWC China Lake data, although the two laboratories were rated separately. This had the effect of increasing China Lake elements such as Total Authorized Personnel and Total Program Dollars by more than twenty percent; other elements were increased in roughly the same proportion.

During the revalidation of the fiscal year 1968 data, the data for Corona and for two other Navy laboratories for which fiscal year 1968 data had been lacking became available. This made the fiscal year 1968 data base the only year complete for all laboratories. By this time I had also come to feel that the time frame of the fiscal year 1968 data might be more representative of the ratings than the fiscal year 1969 data. Therefore, most of the correlations shown in the appendices are based on the fiscal year 1968 data.

The reader may wonder why I did not immediately strike an average of the fiscal year 1967, fiscal year 1968, and fiscal year 1969 data and use the average. For one reason, the data did not become available until late in the study. The fiscal year 1967 data was the last to be validated, and this not completely so (see section 3.5). For another reason, I had no idea what the similarities or differences in the three years of data might be, and I did not want to combine them until I had seen the correlations between the peer ratings and each year separately.

ARMY		NAVY		AIR FORCE	
CIVMS	.364 (23)	CIVMS	.886 (18)	CIVMS	.463 (10)
CIVPH	.381 (23)	CIVPH	.791 (17)	CIVPH	.590 (10)
TMAST	.362 (23)	TMAST	.881 (18)	TMAST	.416 (10)
TPHDS	.368 (23)	TPHDS	.793 (17)	TPHDS	.621 (10)
IH6.1	.520 (21)	IH6.1	.690 (18)	IH6.1	.587 (7)
T6.1$	.442 (21)	T6.1$	.713 (18)	T6.1$	.480 (9)
NONRD	.226 (13)	NONRD	.864 (6)	NONRD	.872 (4)
SEQAS	.265 (22)	SEQAS	.792 (18)	SEQAS	.420 (9)
TBACH	.218 (23)	TBACH	.6[illegible] (18)	TBACH	-.452 (10)
SEQIP	.458 (22)	SEQIP	.724 (18)	SEQIP	-.145 (9)
PAPER	.205 (23)	PAPER	.589 (17)	PAPER	.564 (10)
IH1-2	.415 (22)	IH1-2	.707 (18)	IH1-2	.174 (10)
MILND	.443 (9)	EQUIP	.832 (18)	MILPA	-.764 (4)
OWNED	.547 (18)	LSPAC	.776 (18)	IH6.4	-.771 (5)
LEASD	.862 (5)	IHR&D	.831 (18)	OH6.4	-.877 (6)
OHO&M	.534 (10)	MILCN	.874 (5)	DEPOM	-.811 (5)
OTHPR	.649 (5)	IH6.3	.788 (17)	TO&M$	-.811 (5)
OTHOM	.911 (4)	CPROF	.822 (18)	TNDOD	.702 (6)
MILGS	.551 (12)	TPROF	.801 (18)	T6.4$	-.877 (6)
TODOD	.421 (21)	TR&D$	.779 (18)		
ACRES	.458 (20)	TIH$	.774 (18)		
		IH1-3	.794 (18)		
		IH1-4	.808 (18)		
		T61-3	.793 (18)		

FIGURE 5.9
Principal Correlations, FY 1968 Data

The correlations for the three years of data are shown in Appendix F. In order to make the Navy laboratories comparable, the NOLC data has been recombined with that of NWC. The correlations for the major DoD variables over the three years show a high degree of consistency with the tables of the distribution of data base elements according to percent deviation from the mean shown in Appendix C. The values found to be most stable in Section 3.4 are generally among the DoD correlations showing the least annual variation; an exception is the correlation with the number of wageboard employees (.183 in 1967 versus .332 and .315 in 1968 and 1969).

The values of the correlations of some of the principal elements are shown in Figure 5.10. Included also are correlations for these elements using the data from fiscal year 1970; they are generally consistent with those of the three previous years. In all cases shown, the correlation with the average of the three years of data lies in the range of the correlations for the individual years. This need not be true however, since the number being computed is the correlation of the averages, not the average of the correlations. For example, the average value of In-House 6.1-6.4 in the DoD laboratories is higher than the value for any one of the three years.

It might appear that the consistency of the correlations across the fiscal years of the data base would reinforce the reliability of the peer ratings. It should be stressed that this is not the case. Rather, what is being indicated by the elements for which the correlation is relatively consistent is one or more of three conditions: (1) the data element itself is stable over the three years; (2) the correlation is being influenced by an extreme point, or outlier, so that fluctuations of the other laboratories having the property have little effect on its value; or (3) some sort of internal balancing among the amounts of the elements possessed by the various laboratories is compensating for the annual variations. An example of the second type of effect is shown by the data for Total Authorized Personnel (TAPER) and Total Other DoD Appropriations (TODOD) in the Army laboratories. The relative deviations of these elements show the former to be stable and the latter to be unstable:

---

[1] Pre-publication review of these differences indicates there is an error in the FY 67 WGBRD and CLASS datum for one of the Navy laboratories; the corresponding correlations are probably more like those for FY 68 and FY 69.

CORRELATIONS BETWEEN RATINGS AND LAB ELEMENTS
FISCAL YEARS 1967, 1968, 1969, 1970

LABEL	1967	1968	1969	AVG. 67-8-9	1970
ARMY LABORATORIES					
TBACH	.261	.218	.090	.190	.208
TMAST	.347	.362	.287	.334	.316
TPHDS	.350	.368	.372	.366	.387
TPRØF	.299	.254	.161	.238	.164
TAPER	.258	.254	.221	.248	.230
T6.1$	.472	.464	.417	.452	.402
TR+D$	.186	.171	.128	.163	.185
TPRØ$	.440	.215	.325	.342	.238
TØ+M$	-.029	.040	-.154	-.074	-.238
TPGM$	.203	.238	.137	.198	.220
NAVY LABORATORIES					
TBACH	.715	.687	.708	.705	.661
TMAST	.901	.906	.910	.909	.906
TPHDS	.805	.817	.806	.811	.769
TPRØF	.811	.799	.818	.811	.815
TAPER	.664	.674	.669	.669	.665
T6.1$	.705	.738	.755	.735	.721
TR+D$	.727	.762	.830	.777	.834
TPRØ$	-.268	-.145	-.149	-.185	-.216
TØ+M$	-.392	-.350	-.503	-.450	-.543
TPGM$	.448	.496	.455	.470	.499
AIR FORCE LABORATORIES					
TBACH	-.420	-.452	-.264	-.386	-.354
TMAST	.486	.416	.373	.429	.318
TPHDS	.632	.621	.636	.630	.616
TPRØF	.003	-.043	.093	.019	.002
TAPER	-.104	-.074	-.013	-.063	.022
T6.1$	.519	.484	.477	.494	.562
TR+D$	-.357	-.092	.025	-.159	.162
TPRO$	-.162	-.209	-.310	-.232	-.212
TØ+M$	-.380	-.372	-.373	-.376	-.383
TPGM$	-.364	-.157	-.092	-.217	-.062

FIGURE 5.10

	Number of Army Laboratories With Percent Deviations Between			
	0-10	10-25	25-50	50-200
TAPER	17	3	3	0
TODOD	0	1	5	17

while the annual correlations are fairly stable:

	FY67	FY68	FY69	FY 67-68-69
TAPER	.258 (23)	.254 (23)	.221 (23)	.248 (23)
TODOD	.330 (18)	.389 (21)	.361 (21)	.364 (23)

A graph of the range of Other DoD Appropriations over the three years is shown in Figure 5.11; the average value is the point within the ranges. (The intervals shown as points on the vertical axis were simply too short to graph). It can be seen that the one right-hand value exerts much leverage on the correlation; without it, the value of the correlation coefficient would be considerably different, and probably not as regular with time.

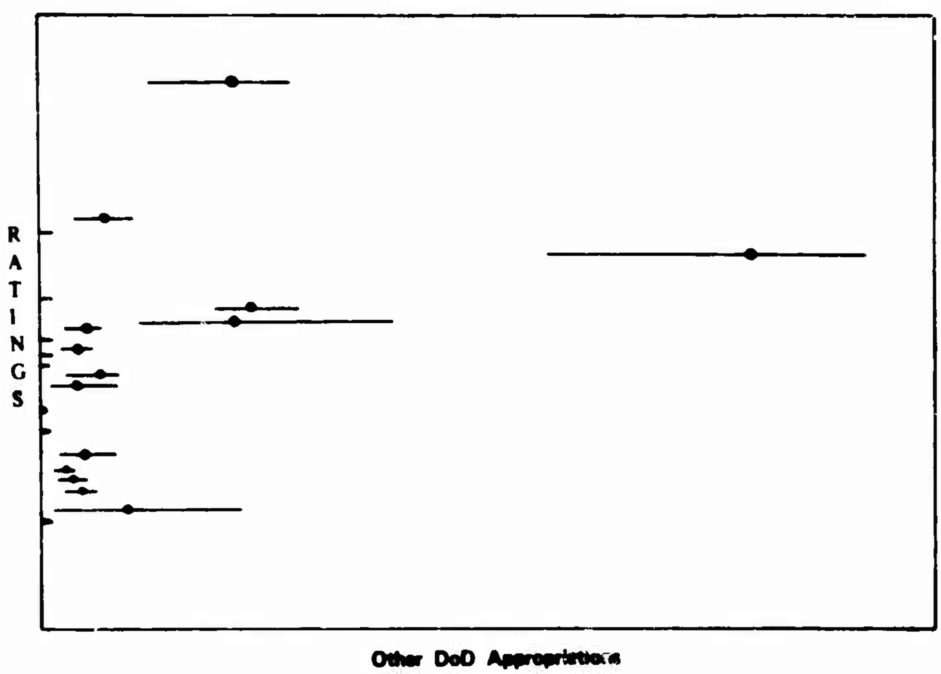


FIGURE 5.11

For the most part, the choice of the fiscal year 1968 data as a basis for the computation of the correlations has been an acceptable compromise. Except for a few instances, the values of the expanded elements generally lie within the interval bounded by the 1967 and 1969 fiscal year data, or within ten percent of one of the endpoints of the interval. However, for the regression analyses and the rank-order comparisons, it would have been better to have used the average of the three years of data, in order to smooth out some of the random fluctuations. This would ignore trends that may be inherent in the data, but there has been no provision to accommodate them anyway.

5.5 Variation by Rater Groups

The correlations between the fiscal year 1968 data elements and the peer ratings were computed using the ratings obtained from a number of different rater groups. In addition to the standard ratings (of all raters combined), these consisted of:

(1) Program Managers in the Office of the Director of Defense Research and Engineering, and on the Headquarters Staffs and Service Commands of the three military departments.

(2) DoD laboratory managers (mainly Technical Directors and Commanding Officers)

(3) Those segments of groups (1) and (2) which could be identified according to military departments, and were accordingly organized in three groups - Army, Navy, and Air Force. The ratings of each service's laboratories were computed using the rankings from only that service, and

(4) Raters from Private Industry.

The correlations for the various groups are shown in Appendix G. (The correlations for the standard ratings are shown in Appendix E.) The correlations between the Industry ratings and the laboratory properties are considerably different from those of the other groups; this is because the ratings are different from those of the other groups. Two of the laboratories have rank-order differences greater than twenty-five between the Industry ratings and the standard set; two others have rank-order differences greater than twenty; and all told, twenty-five have rank-order differences in excess of ten positions.

CØRRELATIØNS ØF DIFFERENT RATER GRØUPS
FISCAL YEAR 1968

CØDE	DSC	LAB	ANF	DØD	IND	ALL
ARMY LABORATORIES						
TBACH	.298	.192	.209	.248	.145	.218
TMAST	.412	.323	.344	.373	.285	.362
TPHDS	.413	.290	.393	.353	.359	.368
TPRØF	.329	.217	.243	.275	.191	.254
TAPER	.302	.197	.230	.249	.265	.254
T6.1$	.516	.426	.452	.479	.347	.464
TR+D$	.256	.105	.113	.182	.145	.171
TPRØ$	.186	.279	.251	.242	.100	.215
TØ+M$	.058	.060	.137	.054	-.005	.040
TPGM$	.302	.199	.198	.254	.179	.238
NAVY LABORATORIES						
TBACH	.693	.698	.705	.705	.471	.683
TMAST	.868	.878	.882	.890	.623	.881
TPHDS	.734	.792	.747	.778	.749	.799
TPRØF	.798	.808	.827	.815	.592	.801
TAPER	.693	.668	.690	.691	.600	.689
T6.1$	.650	.711	.658	.693	.632	.713
TR+D$	.785	.773	.757	.791	.590	.779
TPRØ$	-.157	-.295	-.233	-.231	-.051	-.222
TØ+M$	-.320	-.299	-.289	-.325	-.279	-.319
TPGM$	.527	.443	.473	.491	.436	.488
AIR FORCE LABORATORIES						
TBACH	-.457	-.266	-.426	-.383	-.643	-.452
TMAST	.293	.509	.304	.395	.180	.416
TPHDS	.619	.674	.665	.652	.184	.621
TPRØF	-.083	.157	-.040	.019	-.419	-.043
TAPER	-.065	.084	-.058	-.003	-.506	-.074
T6.1$	.499	.601	.534	.552	-.093	.484
TR+D$	-.094	.018	-.211	-.045	-.310	-.092
TPRØ$	-.029	-.077	.113	-.049	-.693	-.209
TØ+M$	-.288	-.206	-.199	-.260	-.768	-.372
TPGM$	-.099	-.014	-.140	-.065	-.520	-.157

FIGURE 5.12

The differences in rank-order between the Industry ratings and the standard ratings are undoubtedly in part due to the small sample size. Eighteen of the ratings - ten Army and eight Navy - are based on less than fifteen votes; five of the ratings are based on fewer than ten rankings; and one is as low as three. However, the ratings of the four laboratories having the greatest rank-order differences - including one Air Force laboratory - are based on ten or more votes, the average being eighteen.

The Industry ratings of the Air Force laboratories are based on a minimum of twenty-three votes; therefore they are comparable to the DDR&E and laboratory samples in size. However, the differences in the correlations of the ratings and the laboratory properties between the Industry group and the other raters are even more noticeable in the Air Force laboratories. From this, it may be inferred that the differences between the Industry sample and the other groups observed in the correlations of the Army and Navy laboratories are not just because of the small sample size, but also reflect a different point of view.

The variations in some of the principal elements according to different rater groups are shown in Figure 5.12. The correlations shown under the heading "DSC" are those between the laboratory properties and the ratings based on rankings by DDR/E, Headquarters Staffs, and Service Commands. Similarly the correlation between the elements and the ratings by laboratory rankers are shown in the column labeled "LAB". "ANF" stands for Army, Navy, and Air Force; the entries in this column are based on the ratings of Army laboratories by Army participants, Navy by Navy, and Air Force by Air Force. The correlations under the heading "DOD" are based on the ratings of all DoD participants. This is a considerably larger group than "ANF", since it includes the OSD rankings plus those laboratory's participants not identified as to service. (These have not been included in the Appendices.)

For the Army, the correlations using the ratings of DDR&E, Staffs, and Commands are larger than the correlations obtained from the rankings of Army laboratories by Army raters only, except for Procurement Appropriations and Operations and Maintenance. Similarly the ANF correlations are higher than the LAB correlations except for Procurement Appropriations. The Industry correlations are of about the same magnitude as those of the Laboratory group - higher in some places, lower in others - but generally they tend to weaken the correlations established by the DoD raters.

The Navy correlations show a similar but more pronounced pattern between the correlations obtained from the DoD ratings and those from Industry; these latter values are generally much lower, except for the correlations with sources of funding from outside the Department of the Navy. The Navy pattern differs from the Army's in that the correlations among the DoD subgroups are more nearly of the same magnitude. Also, whereas the ratings of the Laboratory Group correlated more positively with Army Procurement than did those of any of the other groups, they correlated more negatively with Navy Procurement than did those of any of the other groups.

The correlations between the ratings and the properties of the Air Force laboratories show the most variation of any of the three services. Within DoD, there are considerable differences between the DSC and the LAB correlations; and between DoD and Industry, the differences are sharper than in the other military departments. The LAB group generally has the highest correlations; the DSC and ANF groups are about equal; and the Industry correlations are usually the most negative. Some of the latter correlations are so negative that in absolute value they exceed the Laboratory Group correlations, e.g., the absolute value of Industry's correlation with civilian professionals is .534, versus .280 for the Laboratory Group. However, as noted earlier, very few of the Air Force correlations are statistically significant (at the .05 level) because of the relatively small number of laboratories. With only ten laboratories, a change in the datum of any one of them can result in fairly large differences in the correlation. For example the correlation between all raters and Acquisition of Scientific and Engineering Equipment with Project Funds (SEQPR) in ten Air Force laboratories for fiscal year 1968 is -.088; but with one laboratory removed, the correlation jumps to .717. Therefore these relatively large variations must be viewed with caution lest they be given undue emphasis.

5.6 Summary

There are significant correlations between the peer ratings and the properties of the Navy laboratories. The Army laboratories for the most part show only minor correlations between the peer ratings and the data base elements. A few of the Air Force correlations are as large as those found for the Navy, but because of the much smaller number of Air Force laboratories, most of the correlations are not as statistically significant.

In many cases the correlations are quite similar for each of the three years of the data base, and also for the averages of the data over the three year period. The correlations with the data for fiscal year 1968 generally lie between those for fiscal year 1967 and fiscal year 1969, and thus fiscal year 1968 appears to have been a fortuitous choice of a base year.

There are considerable differences in the correlations between the laboratory properties and the ratings based on the ranking of different rater groups, particularly between the raters from industry and the raters from DoD. The correlations based upon the industrial ratings are for the most part lower than those based on the DoD ratings, but the effect on the overall ratings is not as large as it might seem, since the DoD group comprises more than 75% of the sample. On the average, the larger of the correlations based on ratings from all the rater groups are within 5% of the corresponding DoD correlations.

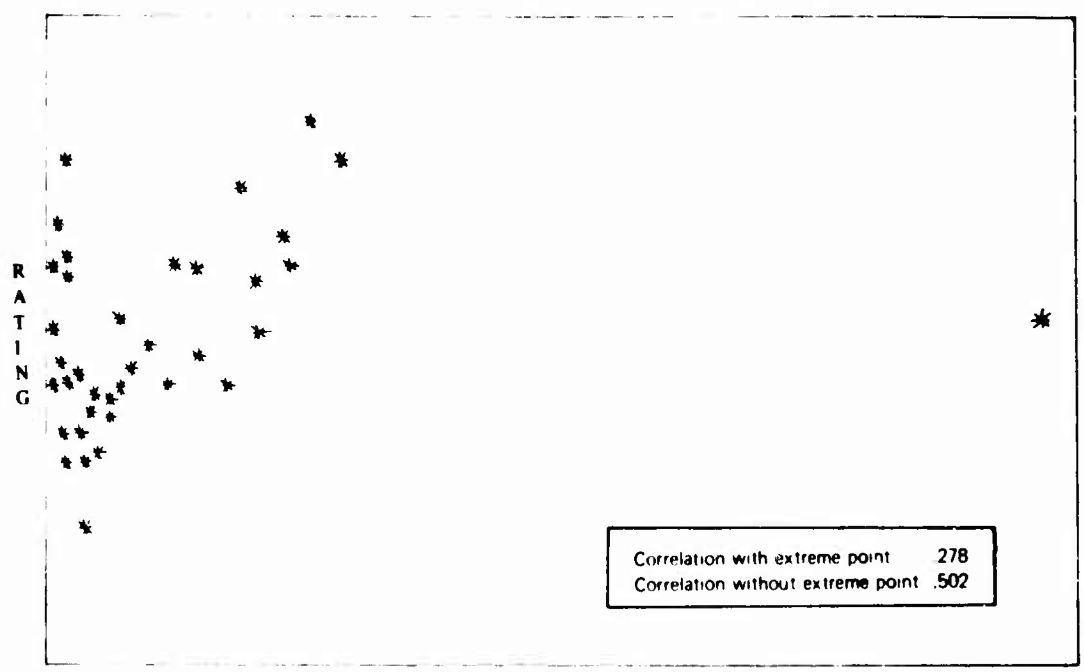
Overall, there is generally more consistency between the two major DoD rater groups than between the DoD and Industry raters. Within the DoD, the two groups (DDR&E, Service Commands, and Service Headquarters; and Laboratories) correlate about equally well with the Navy data and with DoD as a whole, and show opposite trends for the Army and the Air Force data - the DSC ratings correlating higher with the Army, and the LAB ratings correlating more positively with the Air Force.

6. CORRELATION ANALYSIS (II)

6.1 Extreme Points

A prime consideration in the significance and interpretation of the correlations is the distribution of the variables. Two attributes of the distribution of a laboratory's properties were noted in Chapter 3 - density (measured by the proportion of zero to non-zero values), and uniformity (measured by the equality of spacing from high to low). It was also shown that a very few of the laboratories account for a large part of the total of any element. An extreme point, or outlier, vastly separated from the rest of the distribution, can substantially alter the correlation.

The graphs obtained with the correlations produced by the Air Force group at the Pentagon were helpful in pointing up lopsided distributions of the data. An example of the effect of an extreme point is shown in Figure 6.1; this has been taken from reference (3). Using only the non-zero data for thirty-seven of fifty-four DoD laboratories, the correlation of the peer ratings with In-House Advanced Development (IH6.3) is found to be .278; but with the removal of the one out-lying value the correlation is almost doubled, becoming .502.



In-House Advanced Development

FIGURE 6.1

In order to examine the sensitivity of the correlations to the effect of outliers, each element of the fiscal year 1969 data base was examined visually to see if it contained extreme points.[1] An extreme point might be defined as one that is more than three or four standard units (sigmas) from the mean, or one that is the maximum or minimum of a set of points and is two or more times as large (or small) as its nearest neighbor. The trouble with the sigma definition is that the extreme point creates a larger than usual sigma and thus may contain itself. One method of getting around this would be to remove the extreme point, and then test it with respect to the standard deviation of the reduced set of points. However, in the present study, the identification of extreme points was accomplished simply by a visual inspection of the laboratory properties, whereby the largest one or two variables were removed from certain of the elements according to whether or not it was felt they were exceedingly disproportionate. As a rule, a variable was not removed unless it was more than half again larger than the next largest, or unless there were two approximately equal and both much larger (by the same criterion) than the others.

Where the removal of the largest would just cause the next one to pop up with a similar extreme distribution, the item was passed over. Also some items where there were very few non-zero values were passed over. The removals were made with respect to the in-service distribution, not with respect to the DoD as a whole. Therefore, it is valid to make before-and-after comparisons only within a military department. The correlations with the extreme points removed are shown in Appendix G under the column labeled "OUT" (for "Outlier"). Only the basic variables were thus modified; it was not possible to modify the expanded variables due to the mechanism used to remove the extreme points (the value of an outlying point was replaced by -1; an expanded variable thus might be the sum of a real value and a negative 1).

In modifying the data base to suppress the outlying values, it was apparent that not only do a few laboratories account for a large part of the total, but frequently they are the same few laboratories for each of the elements. It was therefore decided to carry out a similar modification, suppressing from each service the laboratory that had the largest total program (this also corresponded to the one having the most people, as shown in Figure 3.13). The correlations under these circumstances are shown in Appendix G under the column labeled "BIG".

The two procedures described above were designed to test the sensitivity of the correlations to large anomalies in the data. A third procedure was also introduced to

[1] The fiscal year 1968 data had not been validated at the time these experiments were conducted.

examine the effect of similar anomalies in the ratings, since the highest rated laboratory in each service was considerably removed from the majority of the others. These correlations are also shown in Appendix H, under the column headed "TOP".

Selected values of the different correlations obtained for each of the three military departments are shown in Figure 6.2; these have been taken from Appendices E, H, and J (Skip Zeros) (In addition to the regular and modified correlations, the correlation obtained using the logarithm of the data base element has been included for comparison under the column headed "LOG". The logarithmic transformation itself is discussed later.)

The various entries in the table shown in Figure 6.2 are

(1) N,R: N represents the number of laboratories having non-zero values of the element; R represents the number of laboratories "removed" in accordance with the visual inspection for outliers.

(2) REG: This is the correlation computed in the "regular" way, using the standard ratings for the N laboratories.

(3) OUT: This is the correlation computed as above, except that R "outliers" have been removed: only (N-R) of the laboratories have been used.

(4) BIG: This is the same as REG, except that the laboratory considered to be the "biggest" in each service has been excluded. In some cases, the correlation is the same as in REG, indicating that in these cases the "biggest" was not the biggest, since it had none of that particular property.

(5) TOP: This is the same as REG except that the highest-rated laboratory has been excluded. As above, it may have had none of the particular property, in which case the correlation is the same as in (2).

(6) LOG: The correlation has been computed using $\log_{10}$ of the laboratory property.

The results shown in Figure 6.2 typify the worst cases found in the conduct of the various tests, i.e., the cases of the more extreme variations. There are several different classes of variation: (1) there is an outlier, other than the BIG or TOP laboratory, exerting considerable leverage upon the correlations (e.g., Army - MILPH); (2) there is an outlier, but it is either the BIG or the TOP (e.g., Army - OH6.3); (3) the visual inspection failed to identify BIG or TOP in a situation where the correlation might change considerably without it (e.g., Air Force - LSPAC); and (4) the removal of BIG or TOP sometimes has quite opposite effects (e.g., Army - OHPRO; Navy - OHOMA; Air Force - TACIV).

Dependency on Large or Higher-Ranked Laboratories

While the correlations shown in Figure 6.2 were selected to show the large variation that may result from the presence or absence of one or two laboratories; they also indicate to some extent how the correlations may be dependent upon either the BIG or the TOP laboratory. Further insight to the effect of these laboratories upon the correlations can be obtained by looking at the correlations of the expanded variables in Appendix G. For the Army, the highest correlation in the expanded variables - total funding from non-DoD sources (TNDOD) - drops from .569 to .280 with the removal of the laboratory rated first; similarly, the correlations with respect to Procurement and Operations and Maintenance are seen to depend very much upon the top-rated laboratory. The Navy expanded variables, on the other hand, indicate more dependency upon the BIG laboratory, particularly with respect to Procurement Appropriations and to all categories of Out-of-House Appropriations. In the case of the Air Force expanded elements, the situation is ambiguous; for most of the elements, the removal of the TOP laboratory sharply reduces the correlations - but on the other hand, the removal of the BIG laboratory changes the correlations about an equal amount in the upward direction. It looks like the TOP laboratory is supporting Total Research Dollars (T6.1$) and the BIG laboratory is depressing Out-of-House Research and Exploratory Development Appropriations; but the patterns in the basic variables (IH6.1, IH6.2, OH6.1, OH6.2) do not bear this out.

Higher-Valued Correlations

Some of the higher-valued correlations that show the most consistency across the various tests are shown in Figure 6.3 for each of the military departments. The Navy correlations are significant at the .01 level or better; the Air Force values are

VARIATIØN IN CØRRELATIØNS DUE TØ EXCEPTIØNAL ELEMENTS
FISCAL YEAR 1969
(SKIP ZERØS)

LABEL	N,R	REG	ØUT	BIG	TØP	LØG
ARMY						
MILPH	17,1	.081	.402	.074	.073	.255
IH6.5	20,1	-.042	-.355	-.355	.041	-.173
ØH6.3	14,1	-.074	-.534	-.534	.043	-.444
ØHPRØ	13,0	.415	.415	.441	-.310	.259
ØHØ+M	13,0	.249	.249	.249	-.374	.105
ØHØMA	6,1	-.572	.759	-.572	-.754	-.491
NØNRD	14,2	.485	-.179	.500	.225	.166
DEPMS	13,1	-.392	.140	-.392	-.452	-.091
DEPPR	18,2	.338	-.241	.346	-.306	.018
NAVY						
TAMIL	17,1	.388	.135	.135	.504	.375
SEQPR	15,1	.759	.441	.717	.641	.682
ØHR+D	17,1	.516	.363	.363	.611	.493
ØH6.1	11,2	.696	.303	.674	.595	.633
IH6.1	16,1	.733	.823	.781	.823	.853
IH6.2	17,1	.822	.787	.824	.787	.885
ØHØMA	12,1	.524	-.235	-.235	.524	.254
ØTHRD	10,1	.627	.371	.593	.371	.363
DEPMS	17,1	.331	-.173	-.173	.444	.207
AIR FORCE						
TACIV	10,2	.015	.064	.417	-.316	.043
CIVMS	10,1	.402	-.051	.452	-.051	.304
LSPAC	10,0	.084	.084	.424	-.066	.385
SEQIP	10,1	-.143	.530	.530	-.249	.141
SECNF	9,0	.096	.096	.356	-.453	-.173
SEQPR	10,1	-.088	.717	.717	-.098	.419
IH6.1	8,1	.568	.490	.548	.490	.549
ØH6.4	7,1	-.833	-.473	-.910	-.841	-.853
HØUSE	8,0	-.163	-.163	.290	-.455	-.228

FIGURE 6.2

significant at the .10 level; except that CIVSV is at the .15 level; an o of the Arm correlations (IH6.1 and OWNED) are significant at the .05 level, two others (CIVPR and SEQPR) are significant at the .10 level, and the fifth is at the .15 level. However, land OWNED is not ideally representative of the Army laboratories since the highest-rated laboratory had none. A similar remark applies to Non-DoD Miscellaneous Source of Funding (NONMS) in the Navy. (It was noted earlier that the correlation of NONMS for all Naval laboratories, including zeros, was .584; this is still significant but at the .05 level.) The largest Air Force correlation (in terms of the absolute value) was the value of -.833 for Out-of-House Engineering Development (OH6.4); however this dropped to -.473 in the outlier test, so it was not included in the values shown in Figure 6.3. (The correlation between the peer ratings and this element using the fiscal year 1970 data (COUNT ZEROS) is -.279; the previous three years it had averaged -.860).

Comparison With Logarithmic Correlations

I had expected to find a fairly strong association between the modified and the logarithmic correlations, based on the supposition that the logarithm of an extreme point would substantially reduce its effect relative to the other variables. The comparison of the logarithmic correlation with the others in Figure 6.2 shows some tendency in this direction, more so with the Army than with the other two services. For example, in six of the seven cases in the Army elements where an outlier was identified as an extreme value, the logarithmic correlations lie generally about mid-way between the regular and the outlier correlations - perhaps thus signaling the presence of an influential outlier. In the seventh case, Out-of-House Other Miscellaneous Appropriations (OHOMA), the pattern does not hold; but then, this is a somewhat sparse element - there are only six non-zero laboratories. (Generally, the sparse or medium dense variables show the greatest amount of fluctuation; for example, the value of OHOMA for the Army ranges from -.754 without the top laboratory to .759 with the top laboratory - but this is simply spurious variation caused by the small sample and the presence of an extreme point.)

However, the logarithmic correlations did not work quite as well for the Navy and Air Force extreme values. For the Navy values, in five cases out of seven (excluding IH6.1 and IH6.2, which were not really presented to illustrate extreme points), the correlations using the logarithms of the data were approximately of the same magnitude as those of the untransformed data. The two cases where there were signals - OHOMA and OTHRD - were for elements that were only medium dense. In the Air Force, the logarithmic correlations approached the midway point

	REG	OUT	BIG	TOP
Army				
CIVPH	.346	---	.353	.488
OWNED	.525	.431	.559	.525
SEQPR	.406	.543	.543	.396
IH6.1	.509	---	.615	.466
OTHRD	.363	.380	.351	.429
Navy				
CIVMS	.911	---	.895	.872
CIVPH	.806	.845	.804	.845
IHR&D	.894	---	.872	.854
NONMS	.913	---	.872	.913
IH6.3	.852	---	.822	.853
Air Force				
CIVPH	.601	.560	.588	.560
IH6.1	.568	.490	.548	.490
PAPER	.596	---	.579	.444
MILGS	-.522	-.433	-.433	-.541
CIVSV	.477	---	.739	.314

FIGURE 6.3
Consistent Higher-Valued Correlations
[Skip Zeros]

in three out of the five cases involving outliers (again, excludin⌐ IH6.1), but for one of these three was hardly any difference between the .odified and regular correlations to start with.

Overall, these various examinations of the possible effects of outliers indicate that, at least for the Army and the Navy, there are relatively few cases where an extreme point unduly causes a highly significant correlation or (2) masks out significant correlations in the remaining variables. The marginal number of Air Force laboratories precludes making a similar statement, one way or the other, about the effect of extrema on their correlations.

6.2 Maxi- and Mini-Laboratories

The preceding analyses were conducted in order to examine the effect of extreme points upon the correlations between the peer ratings and the laboratory properties. In the course of examining the effects of very large values of the laboratory properties, it was observed that a few laboratories were repeatedly the ones with the largest values; these laboratories were the ones identified as the BIG laboratories. Similarly, in the discussion of zeros versus non-zeros, the focus was upon the effect of extremely small (zero) values of the laboratory properties. While it was not especially noted at the time, a number of the smaller laboratories had already been identified and removed from consideration by virtue of the exclusion of the medical and personnel research laboratories.

One might classify the more extreme of these maxi- and mini-laboratories as being atypically large or atypically small according to certain of their characteristics. For example, a maxi-laboratory might be defined as one with more than 800 professionals or an in-house RDT&E program in excess of 40 million dollars. This definition would encompass the six most right-hand laboratories depicted in Figure 3.12. These four Navy and two Army laboratories together comprise more than fifty percent of the DoD laboratory professionals and more than fifty percent of the DoD laboratories RDT&E program.

In a gross sense, the medical laboratories were considered atypical from the non-medical laboratories for reasons previously described these were set outside at the outset of the study. Another group of laboratories, those listed in Figure 6.4 as having "exceptional characteristics", were identified as being atypically small in terms of criteria such as number of professionals, total program in dollars, in-house RDT&E, out-of-house RDT&E, etc.

(Non-Medical Laboratories - FY 69)

Lab Label	Number of Professionals Less than 100	Total Program Less than $5M	In-House[1] R&D Less than $3M	Out-of-House[2] R&D Less than $2M	Zero Dollars[3] In O&M and Procurement	Funded Only[4] From own Department
Army (1)	77	2784	1728	1020	YES	YES
*(2)	69			None	YES	
(3)	29					
(4)	81	2527	739	1542		
(5)		4105		249	YES	
(6)	85	3263	2266	850		YES
(7)	82	2801	2275	392		
(8)				814		YES
Navy *(1)		2994	2994	None	YES	
*(2)		3236	2577	None		
Air (1)	24	532	268	None	YES	YES
Force*(2)		4185	1500	1841	YES	YES

FIGURE 6.4
Exceptional Characteristics

[1] One other Air Force laboratory had In-House R&D less than $3 million.

[2] Two other Army laboratories and three other Navy laboratories were less than $2 million.

[3] Eight of the twelve Air Force labs had neither O&M nor procurement appropriations.

[4] One Navy and one Air Force lab also in this category.

* Army Behavioral Sciences Research Laboratory
Naval Personnel Research Activity
Navy Personnel Research and Development Laboratory
Air Force Human Resources Laboratory

The data shown in Figure 6.4 were compiled while trying to identify certain common characteristics of the mini-laboratories. A number of laboratories had one such characteristic, but only those laboratories with less than 100 professionals or which had two or more of the characteristics were included. The four laboratories marked with an asterisk were found to have principal aspects of their mission in common; they were primarily oriented in the areas of behavioral sciences and utilization of human resources. It was subsequently decided to exclude these laboratories from the general study because they were different in mission, composition, and staffing from the physical sciences and engineering laboratories. The other Air Force laboratory listed in Figure 6.4 is the Air Force Academy's Frank J. Seiler Laboratory; this laboratory was ultimately omitted from the study because of its small size and other atypical characteristics.[1]

The seven remaining laboratories thus identified in Figure 6.4 were all Army laboratories. The smallest of these - in terms of total professionals - was the Aerospace Research Laboratory. I considered omitting this laboratory from the study because of its atypical characteristics, and in fact, in various portions of the study - in some of the regression analyses - it was omitted. However, in the final analyses, I left it in partly because there were already so many other anomalies about the Army laboratories anyway, and partly because the issue of whether to take it out or not was clouded by its antithetical characteristics - it was the smallest in number of professionals, but one of the more highly rated of the Army laboratories; and although it reported no scientific and engineering equipment, its co-location at NASA's Ames Laboratory gave it direct access to an abundance of laboratory equipment.

In looking further into the effect of this laboratory, I decided to conduct some experiments with the other Army laboratories as well. I had already observed the result of omitting the BIG Army laboratory; so similarly I experimented with omitting the SMALL Army laboratory. In parallel with this, a number of other combinations were also tested, as follows:

(1) SMALL:	The Army laboratory with the least number of professionals was omitted
(2) MINMAX:	Army BIG and Army SMALL were both omitted
(3) TWO MINMAX:	The two largest and the two smallest Army laboratories were omitted

[1] This paragraph is repeated from Chapter I in order to provide continuity of expression.

(4) LEAST NINE: The nine Army laboratories with less than 200 professionals each were omitted

(5) POTPOURRI: Six laboratories that seemed at odds with the correlations were omitted

The results of (1), (2), (3), and (5) are shown in Appendix H, using the fiscal year 1969 data and including zeros. The correlations without the nine smallest Army laboratories were computed using the average of fiscal years 1967-68-69. The effects of the variations are most noticeable in the five elements shown in Figure 6.5. The correlations without the least nine are substantially lower than any of the others, indicating a considerable amount of scatter among the larger laboratories.

	ALL	BIG	SMALL	MINMAX	DOUBLE MINMAX	LEAST NINE	POTPOURRI
CIVMS	.288	.360	.345	.448	.549	.315	.721
CIVPH	.378	.386	.446	.457	.420	.384	.529
SEQPR	.380	.493	.434	.567	.680	.236	.810
1H6.1	.510	.611	.515	.608	.618	.489	.727
TNDOD	.543	.554	.602	.617	.596	.474	.653

FIGURE 6.5
Variations in Correlations Between Peer Ratings and Army Laboratories

The correlations in the final column begin to approach those found in the Navy data for a similar number of laboratories. I have not attempted to identify the particular characteristics of the deleted laboratories - I simply omitted some of those that did not "look right". A possibly interesting experiment would be to use sixteen Army laboratories, sixteen Navy laboratories, and the sixteen Air Force laboratories to see how the three services would compare on an equal-numbers-of-laboratories basis (this would require using the Air Force medical and non-medical laboratories together).

6.3 Transformation of Variables

The distribution of laboratory properties among the various laboratories, as discussed in Chapter 3, suggested that a logarithmic transformation of the data might be worthwile. This was tried two ways:

(1) log (rating) vs log (data)
(2) rating vs log (data)

There was relatively little difference between the correlations obtained in these two ways, although both were considerably different from the correlations obtained using the untransformed data elements. A possible reason for the similarity between the two logarithmic transformations is that the normal and the transformed distributions of the peer ratings tend to have pretty much the same shape, as shown in Figure 6.6. The distribution of the standard ratings are on the left, grouped using a class interval of .60; the logarithms of the ratings are on the right, grouped in class intervals of .05. The untransformed ratings are skewed somewhat to the right; the logarithms of the ratings are more centralized.

The correlations between some of the principal elements for fiscal year 1969 and the DoD laboratories are shown in Figure 6.7. These have been taken from Appendix I. The headings RHO_L, RHO, and RHO_U represent the lower value of the 95% confidence interval, the correlation coefficient, and the upper value of the 95% confidence interval. Of the two logarithmic transformations - one with ratings and elements transformed, the other with just the elements - it was decided for any subsequent logarithmic transformations to use the one with just the elements transformed, since this combination gave slightly higher correlations than using logarithms of both the ratings and the laboratory properties.

The correlations presented in Appendix I show the effect of using the two transformations on the basic elements of the data for fiscal year 1969. The correlations on the first four pages are for the regular ratings and regular data; those on the next four pages are for regular ratings and logarithmic data, those on the last four pages are for logarithm of ratings versus logarithms of data. The correlations were computed using the 95% confidence interval program described earlier, but for a slightly different set of Army and Air Force laboratories from those used elsewhere in this report. The Army set included the Behavioral Sciences

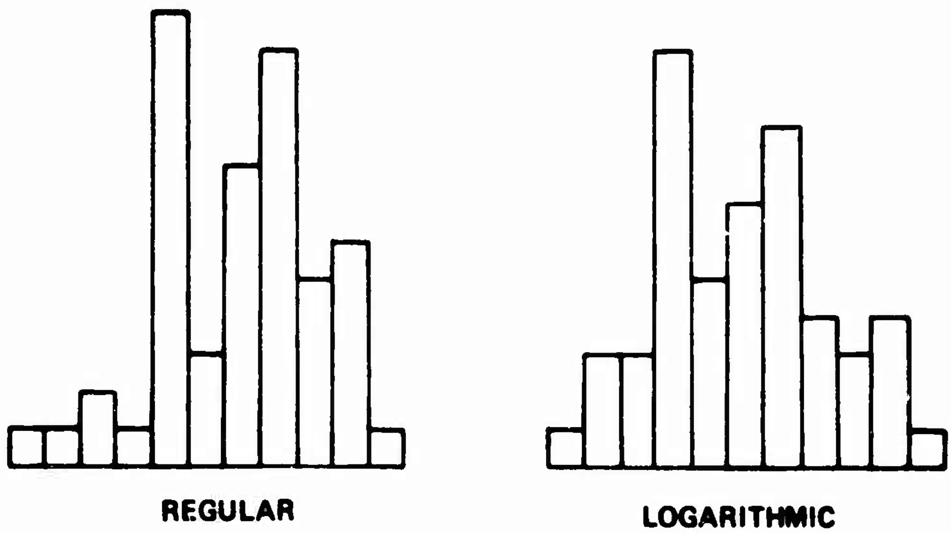


FIGURE 6.6
Distribution of Peer Ratings

	Rating vs Data			Rating vs Log Data			Log Rating vs Log Data		
	RHO_L	RHO	RHO_U	RHO_L	RHO	RHO_U	RHO_L	RHO	RHO_U
CIVBS	.105	.373	.586	.095	.364	.579	.078	.349	.567
CIVMS	.340	.566	.726	.229	.478	.663	.200	.454	.646
CIVPH	.367	.592	.746	.290	.532	.705	.263	.511	.690
EQUIP	.233	.482	.666	.302	.537	.705	.289	.527	.698
IHR&D	.253	.498	.677	.268	.510	.686	.251	.496	.676
OHR&D	.009	.286	.519	.087	.358	.574	.072	.344	.564

FIGURE 6.7
Effects of Logarithmic Transformations
on Correlations for DoD Laboratories

Laboratory but excluded the Aerospace Reserach Laboratory. The Air Force set, in addition to the ten physical sciences and engineering laboratories, included the Human Resources Laboratory. The Navy set was the same. Small variations between these Navy correlations and other fiscal year 1969 Navy correlations are caused by a different order of computation in the programs; the few larger variations are probably attributable to changes in the data elements (these were still being re-validated at the time the confidence interval correlations were computed).

The correlations between the peer ratings and the logarithmically transformed data for fiscal year 1968 are shown in the tables labeled "SKIP ZEROS" contained in Appendix J. (The tables labeled "COUNT ZEROS" were computed by adding 1 to the value of each datum point prior to computing the logarithm; the effect of this latter transformation would be noticeable mainly where the values of the elements are close to zero, i.e., less than 10. The "COUNT ZERO" values were computed preliminary to conducting a regression analysis using the logarithmic data.)

The correlations using the transformed data are generally of the same magnitude as those obtained with the non-transformed data. There is a tendency for the logarithmic values to be somewhat lesser correlated with the peer ratings, except for some of the appropriations for sub-categories of research and development, and certain other elements. Thus the correlations of the major staffing and appropriation elements tend to decrease, while the correlations for the number of PhD's, the value of scientific equipment acquisition, and the research appropriations tend to increase, as shown in Figure 6.8.

The more striking effect of the logarithmic transformation is in the pattern of the distributions. These are depicted in Figure 6.9 for the same elements shown in Figure 6.8. (The distribution of the untransformed data is shown in the upper histogram of each pair; the logarithmic distribution is in the lower.) Note that the change in the distributions of the four elements at the top of the figure, for which the correlations showed a slight decrease, is considerably less drastic than those in the four at the bottom of the figure, in which some of the correlations exhibited a sharp increase.

6.4 Normalization of Variables

One of the first questions that arises in trying to associate the resource data with the peer ratings is: what about size? Is it fair to compare the ratings of one laboratory with a large program and many resources to that of another laboratory with a small program and few resources? What is the relationship between rank and size? Is there a way to normalize the data so that comparisons can be made independently of size?

		Army	Navy	Air Force	DoD
TPROF	Reg	.254	.801	-.043	.436
	Log	.191	.736	-.086	.347
TAPER	Reg	.254	.689	-.074	.399
	Log	.158	.553	-.055	.248
TR&D$	Reg	.171	.779	-.092	.365
	Log	.156	.762	-.158	.402
TPGM$	Reg	.238	.488	-.157	.315
	Log	.158	.376	-.144	.281
TPHDS	Reg	.368	.793	.621	.569
	Log	.424	.846	.555	.542
SEQAS	Reg	.265	.792	.420	.569
	Log	.400	.801	.507	.581
IH6.1	Reg	.520	.690	.587	.578
	Log	.528	.857	.738	.669
OH6.1	Reg	.310	.427	.297	.312
	Log	.522	.486	.520	.546

FIGURE 6.8
Changes in Service Correlation Effected by Logarithmic Transformation

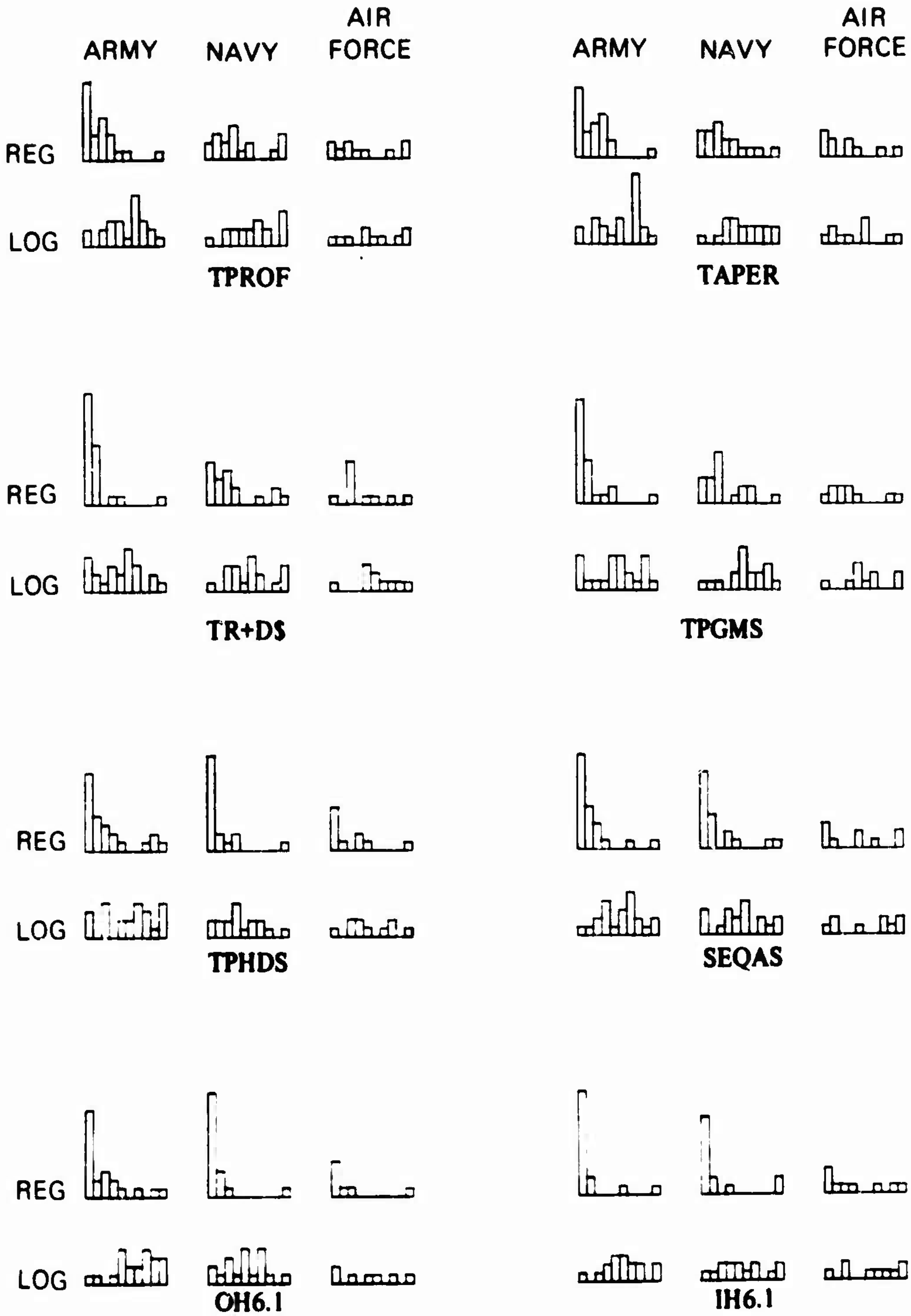


FIGURE 6.9
Regular and Logarithmic Distributions of Selected Variables in the Army, Navy and Air Force

The answers to these questions depend upon what is meant by "normalization", as well as upon the variables being normalized or being used as normalizers. Webster defines normalize to mean "to make normal; to bring into conformity with a standard, pattern, model, etc." In this report, it is used in the context of ratios and proportions, and always involves dividing the values of one element by the corresponding values of another. It is essentially the same process that earlier investigators applied to the data for fiscal year 1968, as described in Section 4.2, except that here it involves a larger set of variables and has been applied to the mean value of the laboratory properties over the three year period consisting of fiscal years 1967, 1968 and 1969. Also, the correlations between the normalized elements and the peer ratings have been computed for each of the three services as well as for DoD as a whole.

If a number of laboratories had equal proportions of a resource, the correlation between their normalized value and their peer ratings would be zero, i.e., the property would have no discrimination power. This is illustrated by Case (1) of Figure 6.10. Element (a) might be the number of patents applied for per year; Element (b) might be the advanced development program in millions of dollars; either (a) or (b) alone might be highly correlated with the peer ratings, but their ratio has no power of discrimination, since it is the same for all. In Case (2) the ratio shows linear discrimination, whereas Element (a) was quadratic and Element (b) was linear. In Case (3), the ratio will be more correlated with the ratings than either of the two elements (assuming the ratings are linear and the elements are listed in rank-order). From these examples, it can be seen that there may be cases where the normalizations increase the magnitude of the correlations, and others where it decreases it.

(1)			(2)			(3)		
Elements		Ratio	Elements		Ratio	Elements		Ratio
(a)	(b)	a/b	(a)	(b)	a/b	(a)	(b)	a/b
50	10	5	45	9	5	20	4	5
40	8	5	28	7	4	16	4	4
30	6	5	15	5	3	24	8	3
20	4	5	6	3	2	12	6	2
10	2	5	1	1	1	8	8	1

FIGURE 6.10
Examples of Ratio Variables

In considering which variables to use as normalizers, i.e., the divisors, the most natural divisors seemed to be the number of professionals employed at a laboratory, the total number of people, the size of the R&D program, and the size of the total laboratory program. Some of these are also elements to be normalized, viz.,

TR+D$/TPROF	Research dollars per professional
TPGM$/TAPER	Total bucks per person
TPROF/TAPER	Professional proportion of total staff
TR+D$/TPGM$	R&D proportion of total program

Do the higher-rated laboratories tend to have more or less research dollars per professional? Or do they have a larger proportion of their budget in R&D$? Or cutting it finer: how does the proportion of PhD's per professional, or the proportion of research dollars per R&D dollar, vary from the higher-rated laboratories to the lower-rated ones? Part of the answers will be considered here, via correlations between such ratios and the peer ratings; another aspect will be presented in Chapter 7.

The correlations normalized by the total number of professionals are shown in Appendix K, for zeros and non-zeros, and including the distribution of the ratios in deci-partitions. It should be noted that the M/M ratio[1] is in several cases close to 1, hence minor changes in proportions could cause substantial changes in the correlations.

Some of the more basic ratios are shown in Figure 6.12. The first ten have been taken from the COUNT ZEROS tables of Appendix K. Since the ratio process could potentially give the same ratio for all laboratories, the correlations are particularly sensitive to the range of the distribution of the ratio values. The deci-partitions of the distribution and the M/M ratios have also been presented in Figure 6.12.[2]

Generally speaking, the ratio correlations are less than those obtained from the simple un-normalized variables. There is a pronounced difference between the correlation of the peer ratings with bachelors per professional [negative] and masters per professional [positive]. However, the M/M ratio is less than two, hence the

[1]This is the ratio of the maximum-to-the-minimum values of the distribution.
[2]If the minimum value is zero, M/M gives the magnitude of the maximum value.

correlations should be viewed with caution. The difference between the correlations of the proportion of research dollars to the RDT&E program versus the proportion of exploratory development is even more pronounced, changing from .55 to -.43 for the Navy laboratories. This result seems suspect at first glance; it seems to say $6.1, Good; $6.2, Bad. But this is contrary to the correlations between the un-normalized values and the peer ratings; in that case, the correlation for T6.1$ was .713, the correlation for T6.2$ was .700. Actually, what the ratio result indicates is that the higher-rated Navy laboratories tend to have a larger proportion of their R&D dollars in research appropriations than do the lower-ranked Navy laboratories; and that the lower-ranked laboratories tend to have a higher proportion of their R&D appropriations in exploratory development dollars than do the higher-ranked laboratories. In terms of the peer ratings, the normalized and un-normalized results are saying: $6.2, Good; $6.1, Better. The situation is depicted graphically in Figure 6.11.

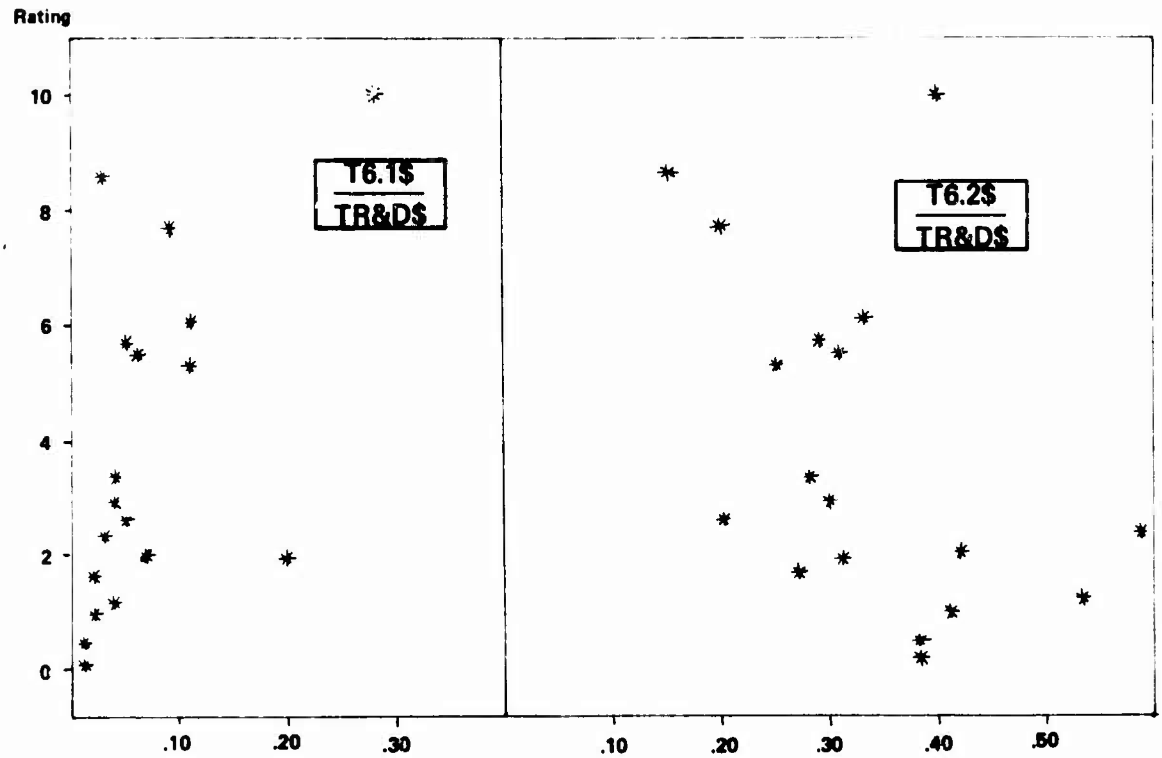


FIGURE 6.11
FY 68 Data, Navy Laboratories

CØRRELATIØNS ØF RATIØ VARIABLES 67-68-69 AVERAGE DATA

RATIØ	ARMY		NAVY		AIR FØRCE		DØD	
TOTND/TPROF	-.381	21	-.361	18	.070	9	-.402	48
TBACH/TPROF	-.266	23	-.174	18	-.668	10	-.287	51
TMAST/TPROF	.474	23	.115	18	.654	10	.408	51
TPHDS/TPROF	.205	23	.421	17	.583	10	.281	50
TR+D$/TPROF	-.060	23	.588	18	-.051	10	.164	51
T6.1$/TPROF	.396	21	.655	18	.514	9	.375	48
EQUIP/TPROF	.092	22	.640	18	.019	10	.293	50
MEETS/TPROF	-.445	23	-.118	18	.359	10	-.057	51
LSPAC/TPROF	.141	23	.479	18	.103	10	.228	51
RPRTS/TPROF	-.249	23	-.292	18	.245	10	-.150	51
TECHS/TAPER	.168	23	.073	18	.346	10	.144	51
WGBRD/TAPER	.149	23	.160	18	.178	9	.106	50
MPROF/TAPER	-.224	23	-.214	18	-.194	10	-.050	51
CPROF/TAPER	.220	23	.027	18	.161	10	.114	51
TPGM$/TAPER	.137	23	-.254	18	-.012	10	.120	51
T6.1$/TR+D$	.387	21	.548	18	.505	9	.339	48
T6.2$/TR+D$	.043	22	-.429	18	-.052	9	.007	49
T6.3$/TR+D$	-.213	17	.166	17	-.201	8	-.090	42
T6.4$/TR+D$	-.231	17	.270	16	-.798	6	-.154	39
T6.5$/TR+D$	-.072	19	-.487	18	.159	6	-.263	43
TR+D$/TPGM$	-.037	23	.514	18	.001	10	.226	51
TPRO$/TPGM$	.043	18	-.419	18	-.039	3	-.204	39
TO+M$/TPGM$	-.108	19	-.380	18	-.343	5	-.253	42
TIH$/TPGM$	.078	23	.395	18	.525	10	.050	51
TOH$/TPGM$	-.078	23	-.395	18	-.525	10	-.050	51

FIGURE 6.12A

DISTRIBUTIØN BY DECI-PARTITIØNS

ARMY	M/M	NAVY	M/M	AIR FØRCE	M/M
34512222-2	1E99	4342121--1	4E 1	2-32----21	7E98
2124224123	1E 0	11-1214413	1E 0	1-1---2213	4E 0
13213722-2	5E 0	1144232--1	5E 0	1-23-1--12	3E 0
53511511-1	2E 1	34431-1--2	2E99	52-1-1---1	2E 1
7444-111-1	8E 0	1244111121	2E 0	133-1----2	4E 0
A52------1	5E 1	76121----1	6E 1	53----1--1	6E 1
1442214122	6E 1	1341114-12	5E 0	12-111-112	1E 1
654223---1	7E 0	474-2----1	5E 1	1-3-2-1-12	1E 1
6942---1-1	4E 1	41413-1211	8E 0	1412--1--1	1E 1
5454112--1	3E 1	37112-2--2	5E 1	8-1------1	2E 1
3444331--1	6E 0	11232422-1	4E 0	242-1----1	2E 1
2354221112	3E 1	1145-411-1	8E 0	331-1---11	2E99
A511-111-1	1E 2	4624---1-1	2E 1	431-1----1	7E 0
1126431-23	3E 0	1--2-26142	3E 0	11--3-21-2	4E 0
84132112-1	6E 0	464111---1	4E 0	412---2--1	4E 0
A422-1---1	7E99	664----1-1	3E 1	71------11	4E99
15243311-3	5E99	2341312--2	4E 0	31-12--2-1	3E99
A432-----1	3E99	12543-11-1	2E99	611---1--1	1E98
A24113-1-1	1E99	91-1--1321	1E99	612------1	9E97
A122-21--1	3E99	A42------1	8E 1	9--------1	1E98
1221-22355	3E 0	2--1-13245	5E 0	2------233	1E 0
A211111-11	6E99	57111--1-2	9E 1	8-----1--1	3E99
A722----11	2E99	61222121-1	3E 1	71-1-----1	6E98
4131221252	4E 0	1-3--42242	5E 0	1121-2--21	1E 1
2521221314	7E99	24224--3-1	1E 1	12--2-1211	1E 0

FIGURF 6.12B

6.5 Combinatorial Products and Ratios

The correlation program was also used to generate all possible combinations of sums, ratios, or products of the data base elements (taken two at a time) using the fiscal year 1967-1968-1969 Averages. Some of the higher correlations between the peer ratings and the combinatorial elements are shown in Figure 6.13. Not all the higher values have been listed; those with larger asymmetrical distributions have not been included.

It was recognized tht some of the combinations would be hard to intepret (i.e., PAPER/TO&M$, etc.), but nevertheless it was decided to look first and explain later. As it turns out, there is not much explaining to do. Most of the correlations were not any higher or any more significant than the simple ones already obtained.

The only Navy correlations that really stood out were the two shown in Figure 6.13. I thought, "Eureka!" when I saw that these also had a generally uniform distribution (i.e., there was no indication of dependency on an extreme point); but on looking into the data further, it turned out that only two of the ten laboratories had more than three military PhD professionals, and that these two had only eight each. The correlations are therefore of little value, since with such extremely small numbers, the addition or subtraction of one person from each laboratory could vastly alter the correlation.

The Army and the Air Force correlations shown in Figure 6.13 are somewhat higher than those found among the unnormalized variables, and while they appear to be substantial, I have not examined them in detail (as was done for the Navy example cited above). One of the Air Force ratios - T61-2/TPGM$ - was used in a regression equation described in Section 8.5. (The other ratio variables used in Chapter 8 were determined prior to the computation of the values shown in Figure 6.13, but were based on a similar computation, with different data.)

6.6 Summary

Experiments using the data for fiscal year 1969 showed that a fairly large number of the correlations changed substantially when outliers were removed. Overall, however, there were relatively few cases among the Army or Navy elements where an extreme point was unduly (1) causing a misleadingly high correlation, or (2) masking out significant correlations in the remaining variables. The marginal number of Air Force laboratories precludes making a similar statement, one way or the other, about the effect of extrema on their correlations.

RATIO AND PRODUCT CORRELATIONS
FY 67-68-69 AVERAGE DATA

VARIABLES	R	N	DENSITY	M/M
ARMY				
I6.1$/MEETS	.639	22	9631--1--1	3E 1
CIVMS/MEETS	.557	23	45144111-2	1E 1
INDOD/T6.3$	.735	15	741111-1-1	3E99
MEETS/IH1-2	-.652	22	7313231--2	1E 1
TOTND/IH1-2	-.637	20	345222-1-3	1E98
INDOD*IH6.1	.637	22	A211--11-1	2E 6
INDOD*MILPH	.604	17	91-22--111	6E 3
NAVY				
TPGMS$/MILPH	.977	10	32-11-11-1	5E 1
TDEP$/MILPH	.976	10	32--11-1-2	8E 1
AIR FORCE				
T61-2/TPGM$	.799	10	1--12112-1	5E 0
OH1-2/TOH$	.766	10	1--3122--1	1E 1
T61-2/T61-4	.800	10	1----223-2	5E 0
TDH$ /T61-3	-.778	10	12121-11-1	3E 0
OH1-3/OH1-4	.835	10	1-------27	2E 0
T61-3/OH1-4	.780	10	1--2122--2	3E 0
MILBS/TPHDS	-.824	10	3-22-1-1-1	1E 2
MILBS/MPROF	-.713	10	11---12122	6E 0
CIVMS/CPROF	.743	10	12-41--1-1	3E 0
PATNT/MILBS	.616	10	5-12-1---1	1E 0
MILPH*OSPAC	.725	10	3-22-1-1-1	2E 1

FIGURE 6.13

Except for a few instances, the Army correlations were not overly dependent on either the largest or the highest-ranked laboratories; the Navy correlations showed a slight dependency on both. The Air Force laboratories were generally higher without the largest laboratory, and lower without the highest-ranked laboratory. Some experiments with a subset of the Army laboratories indicated that in selected circumstances the correlations may be comparable to those found for the Navy variables.

The correlation of the peer ratings with the logarithms of the laboratory properties were generally somewhat smaller than the correlations with the untransformed variables, although the distribution was more centralized and less asymmetrical. The significance of this is probably that the correlations depend primarily on the size of the properties, and the logarithmic transformation considerably reduces the effect of the larger values.

The use of ratio variables - principally the number of professionals and the size of the R&D program substantially reduced the correlation of the Navy variables - again indicating a dependency on size but tended to raise the correlations with the Air Force variables. Some of the more significant of the Air Force ratio correlations appear to result from the large proportion of 6.1 and 6.2 dollars in their appropriations (more than 60%, compared to about 25% each for the other two military departments), and from the relatively large ratio of out-of-house R&D to in-house R&D in the Air Force laboratories (more than 3:1, comparted to 1:1 and 3:7 for the Army and the Navy), but were based on a similar computation, with different data.

7. RANKING ANALYSIS

7.1 Methodology

In the preceding chapters, the association between the peer ratings and the quantitative laboratory properties has been measured according to the correlation between their magnitudes. The purpose of the present section is to examine the peer ratings and the laboratory properties in terms of the rank-order of the ratings, and also in terms of the rank-order of the magnitudes of the laboratory properties. In order to do this conveniently, computer programs were developed to order the data in two different ways - one according to rating and the other according to size.

The ordering by rating depends upon what rating procedure and which rater set is being used; in these analyses, only the standard ratings were used. The magnitude of the properties were listed according to the rank-order of the laboratories; this was a help in recognizing gross characteristics of the laboratories (for example, that one or two laboratories in each of the services tend to account for a large part of the magnitudes of the elements), but the numbers were so large in some elements and small in others, that it made it difficult to make relative comparisons between them. Hence the program was modified to list the percent of each element accounted for by the laboratories. An example of the different perspective one obtains in viewing the element allocations by percent rather than magnitude was shown in Figure 5.6.

An option was also provided to divide all the elements by any one of the elements; this was referred to as "normalizing" the data. This permitted a visual inspection of the laboratory elements in terms of so much per professional, or so much per R&D dollar, or any of a variety of elements. The program also included the capability to sum the percents at specified intervals in the laboratory ranking order (i.e., the first group of five, the second group of six, the third group of four, etc.). This proved especially useful in comparing the higher-ranked laboratories with the lower ranked ones.

The other principal procedure for enumerating the joint distribution of the peer ratings and the laboratory properties was to sort the latter according to size, listing for each laboratory its corresponding rank-order. This permitted visual rank-order comparisons between rank and size, and was used in computing the rank-order correlations shown in Appendix L.

Like the ratings analysis program, the size-effects program has the capability to normalize the laboratory data by any one of the laboratory elements. The program also has a summary feature which permits the compendium of the number of laboratories making up the first xx%, the second yy%, etc., of the service or DoD total. This feature was used to generate the "Percent Accounted for By" data in Table 1 of Appendix C.

The two programs have been quite useful for enumerating certain aspects of the joint distribution of the peer ratings and the laboratory data, but they tend to view the distribution in a one-dimensional way - one does not get the comprehensive impression available from a two-dimensional plot. These latter would provide a simultaneous, concise view of the peer-rating/laboratory-property relationships, such as was shown in Figures 5.1 and 6.1. The programs developed by the Air Force Logistics Support Group at the Pentagon had included scatter-grams as part of the computer output. These had been useful for examining the density (or sparsity) and the uniformity (or non-uniformity) of the distribution of laboratory properties. It was also possible to quickly assimilate which properties had (relatively) little scatter and which had very much.

The capability to view the distribution of the peer ratings versus the laboratory properties was added to the final version of the correlation program described in Section 4.3. The program already showed the deci-distribution of the laboratory properties; this feature was expanded several levels in the vertical direction, and normalized to accommodate the highest- and lowest-rated laboratories in each of the three military departments. Samples of the resulting scatter-grams are illustrated in Figure 7.1. It was intended to use this feature to examine the distribution of selected ratio-variables, but up to this point it has been used only for the unnormalized basic and expanded variables. The diagrams shown in Figure 7.1 excluded the largest laboratory in each service; the correlations (including zeros) were as follows:

	ARMY	NAVY	AIR FORCE
CIVMS	.475	.892	.479
CIVPH	.395	.815	.576
EQUIP	.368	.803	.298
IHR&D	.267	.796	.507
OHR&D	.083	.333	-.057

DISTRIBUTION OF RATINGS (VERTICAL SCALE) AND
FY 68 LABORATORY PROPERTIES (HORIZONTAL SCALE)

```
              ARMY          NAVY          AIR FORCE

CIVMS         -------1--    ---------1    ---------1
              ----------    -------1--    21-1------
              ----1--1--    ----------    1---------
              1------1--    --1-3-----    1--1------
              1-11--1---    ----------    ----------
              21-1-----1    --12------    ----------
              1-2-------    -22-------    ----------
              211--1----    21--1-----    1---------

CIVPH         -1--------    ---------1    ---------1
              ----------    ----1-----    2-11------
              ---1-----1    ----------    1---------
              1--1------    121-------    2---------
              11-1---1--    ----------    ----------
              21-1-1----    3---------    ----------
              2-1-------    3-1-------    ----------
              4-1-------    4---------    1---------

EQUIP         -----1----    ---------1    -----1----
              ----------    ---1------    -11-2-----
              -----1-1--    ----------    -1--------
              1--------1    -11-2-----    ----1----1
              --2-11----    ----------    ----------
              2111------    --11-1----    ----------
              2-------1-    4---------    ----------
              121---1---    22--------    1---------

IHR+D         --1-------    ---------1    ---------1
              ----------    ----1-----    -21-1-----
              ---2------    ----------    ---1------
              1-----1---    --4-------    1--1------
              121-------    ----------    ----------
              3---1----1    -2----1---    ----------
              2-1-------    31--------    ----------
              31-1------    31--------    -1--------

OHR+D         --1-------    --1-------    ---1------
              ----------    -------1--    111------1
              11--------    ----------    -1--------
              1--------1    11--11----    --1--1----
              31--------    ----------    ----------
              22---1----    -1---1---1    ----------
              2---1-----    3-1-------    ----------
              1121------    13--------    --1-------
```

FIGURE 7.1

7.2 Distribution by Rating

There are several different ways of presenting the distributional data obtained by using the programs outlined above. For DoD one might show the percent of a particular characteristic accounted for by quartiles. For the services, the distribution might be shown according to the top-half / bottom-half, top-third / middle-third / bottom-third, etc. For sparse characteristics, the distribution might be shown according to a have/have-not arrangement, i.e., if only six of fifty laboratories have some of a certain property, it might be distributed five among the top half and one in the bottom half. This would provide an intuitive view of which elements were potentially significant in examining associations between the peer ratings and the data base.

Have/Have-Not

Following this latter approach, the listings of the properties for fiscal year 1968 were visually inspected, according to the rank-order of the laboratories, to see how the sparse elements were distributed with respect to the top and bottom laboratories. In most cases, there were very few differences; the most noticeable differences between the upper half of the laboratories in each service versus the lower half are shown below.[†] There were about an equivalent number of elements that were marginally different, i.e., by a value of two or three; but a change of one unit between the two halves would have made them practically indistinguishable on a have/have-not basis.

Army			Navy			Air Force		
	First Half	Second Half		First Half	Second Half		First Half	Second Half
MILPH	11	7	MILPH	7	3	SEQNP	5	1
NONRD	10	3	IH6.6	9	4	OHO+M	0	3
DEPMS	10	6	OH6.6	8	4			
TNDOD	11	5	T6.6S	9	4			

[†]The data for the Aerospace Research Laboratory had been omitted from the data sets used here; therefore the comparisons are for twenty-two Army laboratories.

Upper/Lower Comparisons

In looking at the distribution by quartiles, etc., it was decided, because of the "grey area" in the middle parts of the distribution, and in view of the relative consistency of the ratings within the different military departments (as described in Section 2.5), to utilize only the upper and lower ranked portions of the laboratories. For the DoD laboratories as a group, comparisons have been made using the first seven and the last seven laboratories, the first and last ten, and the first and last thirteen. For the individual military departments the Army comparisons are based on the five highest-ranked and the five lowest-ranked laboratories, the Navy on the top and bottom four, and the Air Force on the first three and last three. Exceptions to these combinations are noted in the text.

Table 1 of Appendix L shows the percent of each element accounted for by the Army, Navy, and Air Force laboratories grouped as described above. Also shown are the values for the high-seven and low-seven, and high-thirteen and low-thirteen laboratories, according to the overall DoD ratings. The values for some of the principal properties are shown in Figure 7.2. These reflect the amount of the various properties apportioned to the high and low laboratories. An estimate of the amounts per professional can be obtained by dividing by the percentages corresponding to TPROF. For example, dividing the percentages shown for TR&DS by the percentages for TPROF yields

ARMY		NAVY		AIR FORCE		DOD	
HI 5	LO 5	HI 4	LO 4	HI 3	LO 3	HI 13	LO 13
1.00	1.00	1.18	.75	.84	1.13	1.03	.87

which indicates that the proportions of TR+DS/TPROF are about equal in the high- and low-ranked Army laboratories, in favor of the high-ranked Navy laboratories, in favor of the low-ranked Air Force laboratories, and slightly in favor of the higher-ranked laboratories on an overall DoD basis.[1]

[1] The ratios are not as exact as they appear to be, since the percents themselves are correct to only one unit. Thus the ratios of the percents for the Army are between 45/39 and 46/38 for the high five, and between 9/13 and 10/12 for the low five.

PERCENTAGES OF ELEMENTS BY HIGH AND LOW GROUPINGS
FISCAL YEAR 1968

	ARMY		NAVY		USAF		DOD	
	HI	LO	HI	LO	HI	LO	HI	LO
	5	5	4	4	3	3	13	13
TBACH	25	14	37	13	17	39	31	16
TMAST	28	11	39	12	36	19	38	12
TPHDS	29	7	59	4	60	7	44	7
TPROF	25	14	38	12	26	31	33	15
TAPER	25	15	40	17	31	33	34	18
T6.1$	40	7	66	2	66	14	59	4
TR+D$	25	14	45	9	22	35	34	13
TPRO$	37	15	24	49	3	58	31	44
TO+M$	21	20	19	35	0	84	22	34
TPGM$	27	14	38	21	20	38	34	20

FIGURE 7.2

PERCENTAGES OF ELEMENTS (LESS WORST CASE)
FISCAL YEAR 1968

	DOD SEVEN				DOD TEN			
	HI	LO	HI -1	LO -1	HI	LO	HI -1	LO -1
TBACH	20	8	15	8	25	11	20	11
TMAST	26	5	20	5	30	7	25	7
TPHDS	35	3	24	3	39	5	28	5
TPROF	22	8	16	8	27	11	23	11
TAPER	25	11	18	10	29	14	22	14
T6.1$	49	2	31	2	57	2	39	2
TR+D$	21	7	16	7	26	9	21	9
TPRO$	26	31	26	18	30	43	30	30
TO+M$	14	24	14	15	16	30	16	21
TPGM$	22	12	16	12	27	16	21	16

FIGURE 7.3

The numbers shown in Figure 7.3 show the percentages of the high-seven versus the low-seven DoD laboratories, and also the high-ten versus low-ten. The entries in the column labeled "HI-1" represent the percentages minus the largest laboratory in the top group; those labeled "LO-1" are minus the smallest in the bottom group (these are reversed for TPRO$ and TO+M$). The intent has been to reduce the effect of the one large laboratory for each case, and to obtain a more conservative comparison between the highs and the lows.

Another way of presenting the data for the same laboratory groupings would be to divide the sum of the elements of the high-ranked laboratories by the sum of the corresponding numbers of professionals and similarly for the low-ranked laboratories. This would give a measure of the average amount of the element per professional. Note that this is not the same as taking the average of the individual proportions of the high group or the low group, but if there are no extreme points in the data, the two results should be in general relative correspondence. For example, the values of TR&D$/TPROF based on dividing the sum of the research and development dollars in the high seven DoD laboratories by the corresponding sum of professionals, and the same for the low seven laboratories, are

	HI 7	LO 7
TR&D$/TPROF	65.8	70.0

whereas when computed by averaging the corresponding individual ratios for the seven laboratories the values would be

	HI 7	LO 7
TR&D$/TPROF	67.8	75.7

There is a larger difference in the low values because one of the lower laboratories had twice as much TR&D$/TPROF as did any of the other lower-ranked ones.

Some of the various properties normalized by dividing the sum of the high or low groups by the corresponding sum of total professionals are shown in Figure 7.4. The values of the "middle" group have also been included for comparative purposes. In more than half of the cases the ratios of the middle group lie between those of the high and low for the group. The high and low groups are the same as in Figure 7.2, except for DoD the comparisons are for the high seven and the low seven laboratories.

PRØPERTIES ØF HIGH-, MIDDLE-, AND LØW-RANKED LABØRATØRIES
PER PRØFESSIØNAL

	ARMY			NAVY		
	HI	MIDL	LØ	HI	MIDL	LØ
TBACH	.650	.662	.681	.657	.688	.709
TMAST	.192	.175	.139	.201	.194	.187
TPHDS	.108	.099	.046	.114	.056	.024
IH6.1	6.9	3.2	2.7	7.8	2.8	.7
OH6.1	1.4	1.2	.1	.5	.4	.1
T6.1$	8.3	4.4	2.8	8.3	3.2	.8
IH6.2	15.2	10.9	8.8	14.3	11.6	13.2
OH6.2	6.3	4.8	10.2	2.6	3.2	3.4
T6.2$	21.5	15.7	19.0	16.9	14.8	16.6
IH1-4	28.0	25.3	18.2	34.8	25.7	18.3
OH1-4	12.4	28.5	36.0	10.6	12.3	5.7
T61-4	40.4	53.8	54.2	45.4	38.0	24.0
IHR+D	35.6	36.2	28.7	46.1	35.8	28.7
OHR+D	33.5	35.3	40.0	18.7	15.8	11.1
TR+D$	69.1	71.5	68.7	64.8	51.6	39.8
SEQNP	.44	.40	.83	1.17	.51	.38
SEQPR	1.30	2.32	.38	2.84	2.36	1.37
SEQAS	1.74	2.72	1.21	4.01	2.87	1.75
PATNT	.068	.039	.058	.057	.034	.058
PAPER	.090	.095	.097	.119	.099	.089
RPRTS	.252	.195	.460	.313	.388	.346

FIGURE 7.4A

PRØPERTIES ØF HIGH-, MIDDLE-, AND LØW-RANKED LABØRATØRIES
PER PRØFESSIØNAL

AIR FØRCE			DØD			
HI	MIDL	LØ	HI	MIDL	LØ	
.406	.626	.781	.604	.674	.660	TBACH
.341	.262	.156	.230	.195	.151	TMAST
.230	.078	.025	.134	.078	.052	TPHDS
25.2	2.5	.6	11.6	3.1	2.5	IH6.1
15.9	4.7	7.4	3.5	1.8	.3	OH6.1
41.1	7.2	8.0	15.1	4.9	2.8	T6.1$
11.1	13.6	20.8	11.8	13.7	9.0	IH6.2
22.7	45.8	56.9	4.3	13.2	10.4	OH6.2
33.8	59.4	77.7	16.1	26.9	19.4	T6.2$
36.6	16.4	22.2	33.9	25.1	17.6	IH1-4
43.0	78.0	92.5	14.4	30.7	28.0	OH1-4
79.6	94.4	114.7	48.3	55.8	45.6	T61-4
37.2	17.8	22.3	42.9	33.4	32.1	IHR+D
49.2	87.6	96.7	22.9	37.8	37.9	OHR+D
86.4	105.4	119.0	65.8	71.4	70.0	TR+D$
1.00	.08	.28	1.14	.38	.68	SEQNP
2.47	2.44	1.18	2.79	2.13	.79	SEQPR
3.47	2.52	1.46	3.93	2.51	1.47	SEQAS
.028	.032	.016	.056	.039	.056	PATNT
.221	.227	.042	.135	.107	.129	PAPER
.186	.345	.089	.283	.289	.118	RPRTS

FIGURE 7.4B

7.3 Distribution By Size

The rating analysis procedure ordered the laboratories according to rating, showing with each laboratory its proportion of the quantitative properties described in the data base. The sort-by-size procedure sorts the elements according to magnitude, associating with the ordered magnitudes the rank of the laboratories.

A measure of the association is given by Spearman's coefficient of rank-order correlation. Denoting the differences between the natural numbers $1,2,\cdots,n$ and the rank of the i^{th} laboratory (ordered by size) as ϵ_i, then the rank-order correlation (RHO) is computed from the equation

$$\text{RHO} = 1 - \frac{6\Sigma\, \epsilon_i^2}{n(n^2 - 1)}$$

where n spans the range of non-zero values of the particular elements. In the event two or more laboratories had the same amount of a roperty, the rank order positions were averaged. The correlations are shown in Table 2 of Appendix L; the correlations for the principal elements are shown in Figure 7.5, together with the correlations obtained by the product-moment method used in Chapter 5. The correlations for the people elements are generally in better agreement than the financial elements, probably because they are somewhat more uniformly distributed, i.e., not quite as clustered (see Figure 3.10). McCloskey notes in reference [9], "in general, if the original data is evenly spread with respect to the two variables, then the rank-order correlation coefficient and the product-moment correlation coefficient will be very close to each other, since little distortion is produced by the transformation to ranked data. If, however, a plot of the original data indicates that the points appear in clusters, it is entirely possible that (the two correlation coefficients) will be considerably different".

High-Low Juxtapositions

By browsing through some of the rank-order listings thus obtained, one can get a feel for the difficulty in finding meaningful relationships between the peer ratings and the laboratory properties. For example, designating the occurrence of either of the first two laboratories as "A", and the occurrence of either of the last two by a "B", then the sequence AB (or BA) occurs thirty-seven times in the sixty basic Air Force elements having three or more non-zero values.[1] The same combinations (1,10; 2,10; 1,9; or 2,9) occur twenty-six times in the thirty-five expanded elements, and

---

[1] The sequence 1,9,2 is counted as only one occurrence, but the sequence 1,9,2,10 is counted as two occurrences.

COMPARISON OF PRODUCT-MOMENT CORRELATION COEFFICIENT (UPPER VALUE) WITH RANK-ORDER CORRELATION COEFFICIENT (LOWER VALUE)

	ARMY	NAVY	AIR FØRCE	DØD
TBACH	.218	.683	-.452	.346
	.214	.703	-.507	.297
TMAST	.362	.881	.416	.577
	.367	.858	.236	.571
TPHDS	.368	.793	.621	.569
	.495	.831	.785	.563
TPROF	.254	.801	-.043	.436
	.291	.808	-.200	.426
TAPER	.254	.689	-.074	.399
	.311	.581	-.030	.276
T6.1$	.442	.713	.480	.525
	.569	.847	.600	.674
TR+D$	.171	.779	-.092	.365
	.175	.761	-.285	.450
TPRO$	.222	-.222	---	-.020
	-.121	-.009	---	.005
TO+M$	.098	-.319	-.811	-.141
	.156	-.117	-.900	-.017
TPGM$	.238	.488	-.157	.315
	.161	.395	-.176	.294

FIGURE 7.5

thirty times in the normalized basic elements basic elements, i.e., those divided by the total number of professionals. There are another nineteen instances wherein only only one laboratory separates what would otherwise be an AB or BA arrangement. Among the Navy laboratories, similar combinations of extreme pairs (1,18; 1,17; 2,18; 2,17) occur eleven times in the unnormalized properties and eleven times in the normalized; and among the Army laboratories, there are thirty-one occasions in sixty-seven elements where A = (1,2, or 3) is adjacent to B = (21,22, or 23), and twenty-nine such occurrences in the normalized basic elements.

High-Low Contrapositions

Another way to examine the data by size and rank is to record the number of times any of the top twenty DoD laboratories appear in the top ten positions of an element, and to similarly record the number of times the bottom twenty appear in the top ten, the number of times the bottom twenty are in the bottom ten, and the number of times any from the top twenty are in the bottom ten. The numbers ten and twenty were arbitrarily chosen; other numbers would of course give different results. Doing this for some of the principal elements gives the data in the upper part of Figure 7.6. The notation T/T,B/T, B/B, T/B stands for Top in Top, Bottom in Top, Bottom in Bottom, and Top in Bottom. Subtracting Column 2 from Column 1 gives a measure of the element's power to associate the higher-ranked laboratories with size; the difference between Columns 3 and 4 provides a similar measure for the lower-ranked laboratories. The product of the two differences, shown in Column 7, may be taken as a measure of the power of the element to discriminate among the rankings of laboratories according to size. The most powerful elements according to these criteria are total research appropriations (T6.1\$), funding from non-DoD sources (TNDOD), the number of professionals with Masters degrees (TMAST), and the combined size of the research and exploratory development appropriations (T61-2).

The same data normalized by the total number of professionals are shown in the lower part of the figure. In this case the principal discriminators are found to be research dollars (T6.1\$), research and exploratory development appropriations (T61-2), equipment (EQUIP), and scientific equipment acquisition (SEQAS).

Since the maximum value that the numbers in Column 7 can achieve is 100, dividing by 100 gives a pseudo correlation coefficient. This is less than either of the coefficients shown in Figure 7.5, but for T6.1\$ the different values are quite comparable.

Unnormalized DoD High-Seven, Low-Seven

	T/T	B/T	B/B	T/B	1 - 2	3 - 4	5 X 6
TBACH	5	0	6	1	5	5	25
TMAST	7	0	8	1	7	7	49
TPHDS	6	2	7	2	4	5	20
TPROF	6	0	7	1	6	6	36
TAPER	5	1	5	3	4	2	8
T6.1$	8	0	8	0	8	8	64
TR+D$	6	0	5	1	6	4	24
TPRO$	3	5	5	4	-2	1	-2
TO+M$	2	5	5	4	-3	1	-3
TPGM$	5	2	5	2	3	3	9
SEQAS	7	1	7	3	6	4	24
EQUIP	7	0	8	2	7	6	42
T61-2	6	0	7	1	6	6	36
TODOD	8	0	8	1	[illegible]	7	56
TNDOD	7	0	9	0	7	9	63

DoD High-Seven, Low-Seven Normalized by TPROF

	T/T	B/T	B/B	T/B	1 - 2	3 - 4	5 X 6
TBACH	1	5	2	6	-4	-4	16
TMAST	6	3	7	0	3	7	21
TPHDS	5	3	7	1	2	6	12
TPROF	-	-	-	-	-	-	-
TAPER	4	4	5	5	0	0	0
T6.1$	8	0	9	0	8	0	72
TR+D$	5	3	6	0	2	6	12
TPRO$	3	6	6	4	-3	2	-6
TO+M$	2	6	5	4	-4	1	-4
TPGM$	4	5	5	1	-1	4	-4
SEQAS	7	2	7	2	5	5	25
EQUIP	6	1	8	0	5	8	40
T61-2	8	1	7	0	7	7	49
TODOD	5	4	7	3	1	4	4
TNDOD	6	2	7	3	4	4	16

FIGURE 7.6
Differential Comparisons Between Rank-Size Effects of Top- and Bottom-Ranked Laboratories

7.4 High-Low Correlations

The correlation program described in Section 4.3 was used to supplement the preceding comparisons of the properties of the high- and low-ranked laboratories. Correlations between the peer ratings and the thirty-five expanded laboratory elements are shown in Figure 7.7. The first column shows the correlations obtained using the first seven and last seven DoD laboratories; the second column shows the correlations using the logarithms of the elements; the third column contains the correlations for the variables normalized by the number professionals, and the fourth column is based on the logarithm of the normalized variables used in Column 3. The variables that show the highest correlations for each of the categories are shown below. The number beneath the category label is the minimum value of the correlation in the set.

HI/LO .810	LOG HI/LO .850	HI/LO NORM TPROF .610	LOG HI/LO NORM TPROF .680
CPROF	TMAST	TMAST	TMAST
TMAST	TPHDS	TOTND	TPHDS
IH1-4	IH1-4	IH1-2	TSPAC
T61-4	SEQAS	IH1-3	TNDOD
TR+D$	EQUIP	IH1-4	T6.1$

The correlations with the expanded variables were also computed for the high and low groups in the separate services.[1] The three highest in each military department were as follows:[2]

ARMY HI 4, LO 4		NAVY HI 4, LO 4		AIR FORCE HI 3, LO 3	
SEQAS	.931	T61-4	.972	MPROF	-.737
TMAST	.885	CPROF	.971	TPHDS	.688
IH1-4	.834	TPROF	.970	OH1-4	-.535

[1] The Aerospace Research Laboratory was omitted from the Army data.

[2] The Air Force values are shown only for elements that are completely dense.

INDEX	CODE	HIGH 7, LOW 7 R	N	LOG HIGH 7, LOW 7 R	N	HI/LO NORM TPROF R	N	LOG HI/LO NORM TPROF R	N
75	MPROF	0.331	14	0.412	14	-.253	14	0.046	14
76	CPROF	0.802	14	0.728	14	0.253	14	0.264	14
77	TPROF	0.820	14	0.783	14	0.000	14	0.000	14
78	TBACH	0.711	14	0.677	14	-.480	14	-.272	14
79	TMAST	0.872	14	0.901	14	0.623	14	0.862	14
80	TPHDS	0.780	14	0.863	14	0.575	14	0.760	14
81	TAPER	0.609	14	0.531	14	-.211	14	-.387	14
82	TSPAC	0.530	14	0.599	14	0.227	14	-.688	14
83	TR+D$	0.815	14	0.740	14	0.039	14	-.408	14
84	TPRC$	0.050	14	0.311	14	-.194	14	0.154	14
85	TO+M$	-.295	14	-.079	14	-.444	14	-.334	14
86	TPGM$	0.604	14	0.533	14	-.186	14	-.544	14
87	TIH$	0.682	14	0.634	14	0.167	14	-.642	14
88	IH1-2	0.717	14	0.857	14	0.617	14	0.203	14
89	IH1-3	0.759	14	0.838	14	0.668	14	0.120	14
90	IH1-4	0.811	14	0.837	14	0.761	14	0.193	14
91	OH1-2	0.253	14	0.393	14	-.096	14	0.131	14
92	OH1-3	0.226	14	0.320	14	-.162	14	0.061	14
93	OH1-4	0.285	14	0.287	14	-.229	14	0.016	14
94	TDEP$	0.587	14	0.490	14	-.222	14	-.471	14
95	TODOD	0.407	14	0.634	14	-.036	14	0.309	14
96	TNDOD	0.517	14	0.772	14	0.533	14	0.689	14
97	T6.1$	0.615	14	0.811	14	0.527	14	0.698	14
98	T6.2$	0.719	14	0.702	14	-.107	14	-.289	14
99	T6.3$	0.368	14	0.376	14	-.199	14	0.165	14
100	T6.4$	0.386	14	0.263	14	-.051	14	0.107	14
101	T6.5$	0.111	14	0.361	14	-.402	14	0.118	14
102	T6.6$	0.576	14	0.471	14	0.461	14	0.381	14
103	T61-2	0.740	14	0.796	14	0.184	14	-.056	14
104	T61-3	0.792	14	0.725	14	0.050	14	-.157	14
105	T61-4	0.828	14	0.693	14	0.035	14	-.205	14
106	TOH$	0.197	14	0.265	14	-.256	14	-.288	14
107	ACRES	0.335	14	0.601	14	0.335	14	0.439	14
108	SECAS	0.705	14	0.871	14	0.621	14	0.536	14
109	TOTND	0.064	14	0.187	14	-.654	14	-.294	14

FIGURE 7.7

7.5 Summary

An examination of the fiscal year 1968 data base elements for each military department, each with the values of each element listed according to the rank-order of the laboratories, showed very few differences on a have/have-not basis between the upper half and the lower half of the distribution. However, when viewed according to size, or the percent portion of the departmental total, the highest-ranked laboratories generally had more than double the amount of the lowest-ranked laboratories. For several variables, such as the number of PhD's and the amount of research dollars, the few top-rated DoD laboratories had six to ten times as much of the property as the few bottom-rated laboratories.

When viewed on a per-professional basis, the ratios for these same variables were about half as large - the highest-rated laboratories having from three to five times the proportion of the lowest-rated ones. There was a large variation in the proportions of the total RDT&E program between the high- and low-rated laboratories of the different services. The Army ratios were about the same - 69.1 thousand dollars per professional in the upper laboratories, 68.7 thousand dollars per professionals in the lower ones. The corresponding Navy and Air Force numbers were 64.8 : 39.8 and 86.4 : 119.0, respectively.

In most cases when the laboratories are ordered according to their proportion of an element, there is at least one instance where a high-rated laboratory is adjacent to a low-rated laboratory (high-low juxtaposition); thus any statements of the sort "the high-rated laboratories have this much, whereas the low-rated laboratories have that much" must be taken advisedly. Generally this has been provided for in this report by phrases such as "the higher-rated laboratories tend to ...".

Another way to examine the data by size is to observe the number of times the low-ranked laboratories appear in the high-order positions, and vice versa (high-low contrapositions). The variables showing the least number of such occurrences, i.e., the variables such that high corresponds to high and low corresponds to low, are Total Research Appropriations, Funding from Non-DoD Sources, the number of Professionals with Masters Degrees, and the combined size of the Research and Exploratory Development Appripriations. When the data is normalized by dividing by the total number of professionals, the variables that most consistently are in the proper position (by the above criteria) are Research Dollars, Research and Exploratory Development Appropriations, Equipment, and Scientific Equipment Acquisition.

8. REGRESSION ANALYSES

8.1 Purpose and Scope

Up to this point we have considered only single-variable associations between the peer ratings and the data base. In the present chapter we shall examine a relationship of the sort

$$R = A_0 + A_1X_1 + A_2X_2 + \cdots + A_nX_n$$

where R represents the rating of a laboratory, the X's stand for the values of certain of its measurable properties, and the A's are coefficients which indicate the extent of the contribution of the various properties to the overall rating. The equation connecting the rating to the properties is called a *regression equation*; the X's are the *independent variables*; R is the *dependent variable*; and the mathematical selection and processing of the variables, together with the associated statistical evaluation of the results, is called *regression analysis*. In the ideal situation, where the properties are independent of one another (i.e., a change in one of the properties does not in itself entail changes in any of the properties), the coefficient A_i determines how much a change in X_i will change R.

In the present case, the situation is far from ideal. The data within groups of elements are highly intra-correlated (e.g., TACIV, CPROF, CLASS, TECHS, CIVBS, etc.), and between the various groups there exists much inter-correlation (e.g., CIVPH, IHR&D, PAPERS, etc.). Hence, a mathematical relationship implying that a laboratory could improve its rating by obtaining more of this element or less of that element would be misleading if just taken at its surface appearance. And from a practical viewpoint, it might be internally or externally unfeasible for a laboratory to follow the model's guidance. If it were true, for example, that a certain proportion of PhD's resulted in an optimum rating, the laboratory's own objectives or those of its parent military department might well preclude the achievement of the optimum proportion. Or even assuming that all laboratories were free to adjust their resources to the proportions indicated by the model, would they then all attain the same rating, or even raise their ratings over the present level? Most likely not, for since the ratings are based on relative rankings, the lower ones that go up must do so at the expense of the ones that go down.

Nevertheless, as Draper and Smith point out in reference [9], "It is in the construction of this type of predictive model that multiple regression techniques have their greatest contribution to make. These problems are usually referred to as *'problems with messy data'* - that is, data in which much intercorrelation exists. The predictive model is not necessarily functional and need not be useful for control purposes. This, of course, does not make it useless, contrary to the opinion of some scientists. If nothing else, it can and does provide guidelines for further experimentation, it pinpoints important variables, and it is a very useful variable screening device."[1,2] It is from this point of view that the present regression analyses have been conducted: to indicate which variables are most significantly related to the peer ratings, to measure the relative strength of the relationship (using the multiple correlation coefficient), and to obtain insight to the rationale used by the participants in ranking the laboratories.

8.2 Methodology

The regression models were used to examine the multivariate relationships between the peer ratings and the quantitative laboratory properties. These were DA-MRCA (the Dahlgren Multiple Regression and Comprehensive Analysis Model, developed at the Naval Weapons Laboratory) and BMD02R (the Stepwise Regression Model developed by the Health Services Computing Facility at UCLA).[3] The MRCA program was used initially to examine the effect of using various blocking variables and to make an initial fit to the fiscal year 1969 data. The results obtained from the use of this program are described in the following paragraphs. The model was applied to the DoD laboratories as a group and also to the laboratories of the three military departments. The BMD02R program was used later in the study, for a different base year (fiscal year 1968), using a different methodology, and for slightly different sets of laboratories and data base elements. Thus only a general sort of comparison can be made between the results of the two models.

[1] From **Applied Regression Analysis** by N. R. Draper and H. Smith. Copyright (C) 1966, John Wiley & Sons, Inc. By permission of John Wiley and Sons, Inc.

[2] I would be remiss in objectivity if I did not note that in the next sentence the authors say: "It is necessary, however, to be very careful in using multiple regression, for it is easily misused and misunderstood."

[3] I am indebted to Ray Brancolini, Gary Gemmill, and Marlin Thomas of the Naval Weapons Laboratory for their help in carrying out these phases of the study. The description of the DA-MRCA analyses was in large part prepared from material supplied by Dr. Thomas and Mr. Gemmill.

Both models start out by selecting the single variable most highly correlated with the peer ratings, and at each step both models select from among candidate variables the one that will contribute most to the total correlation at that step, i.e., considering the inter-correlation with all the other variables in the regression equation. The models differ principally in that BMD02R also at each step examines the variables presently in the regression equation to see if one or more should be taken out. (It is possible that the contribution of earlier variables, selected on a sequentially comparative basis, may have no longer been as significant because of their correlation with later variables.)

Since neither of the two models was able to accommodate the entire list of seventy basic variables, it was necessary to screen out a number of them at the start. For the MRCA runs, this was accomplished by selecting the forty-five variables having the highest correlations with the peer ratings. I had reservations about using some of these, for it was not clear how to interpret them if they were selected, especially the various miscellaneous appropriations and program elements; but I felt at this point I should not over-determine the selection process by too selectively screening the variables. In the BMD02R regression computations, which were limited to forty variables, I omitted plant and facility elements such as OWNED, LEASED, RPROP, EQUIP, and SEQIP, and also omitted the elements having to do with source of funds, in order to concentrate on the remaining basic elements. I had planned to test the omitted properties in subsequent computations, but this has not been done.

However, the major point of difference between the regression computations using the MRCA forward selection process and the BMD02R stepwise procedure is in the way that the peer rankings were utilized. The MRCA program entered the individual rankings of each participant, whereas in the BMD02R program the mean of the rankings, i.e., the standard rating, was used.[1] Thus if one were to compute a simple linear regression of the form

$$R = A_0 + A_1 X$$

as a model of the relation between the peer rankings and property X, the situation would be as depicted in the upper diagram in Figure 7.1. The regression procedure would obtain the equation of the straight line that best fits the data in the least

[1] These are not restrictions inherent in the models, but reflect rather the author's election to use them in this way.

squares se Alternatively, the same sort of model using the average value of the rankings t each laboratory is depicted in the lower diagram. Obviously a much better least squares fit can be obtained by using the mean value, but concomitantly this may inspire a false sense of confidence about the precision of the ratings, since it tends to make them look more exact than they really are.

If all one had were the ratings, i.e., the mean values of the rankings, then these would form an appropriate basis for regression. But since we have the rankings upon which the ratings were based, to ignore them would be to discard pertinent information which indicates the large variation in opinion among the raters; there is no statistical justification for doing this. On the other hand, I intuitively felt that the large variation among the individual raters tended to mask the consistency and stability of the overall ratings; at the 95% confidence level, most of the ratings lie within a one-decile interval - and the general agreement with the Apstein ratings lends credence to the supposition that the ratings are relatively stable.

A compromise between these two points of view might be to use the mean value of the rankings, i.e., the standard ratings, but to weight them according to their respective sample sizes. Using the square root of the sample size as a comparative measure, six of the fifty-one laboratories had a value less than nine (but none less than six), and six had a value greater than thirteen, (but none greater than sixteen); so that by far the majority lay between nine and thirteen. Hence the weighting would for the most part no more than double the importance of the ratings. Also, since the ratings are already significantly correlated with the number of rankings (about 8 for the Navy laboratories), weighting them according to the number of rankings would seem to be attaching more emphasis than they deserved. As it turns out, this part of the discussion is academic, since neither MRCA nor BMD02R had the capability to assign weights to the individual mean rankings, other than to enter the average score for each laboratory according to the number of times it was rated (as was done in MRCA for the individual scores).

To sum it up: in MRCA, the ratings were essentially weighted by the number of rankings, since all the rankings of each rater, (subject to the threshold of ten, as discussed in Chapter 2), were included with each laboratory property; whereas in BMD02R, the ratings were regarded as exact values subject to no variation. Consequently, the correlations using BMD02R are higher than the correlations resulting from MRCA, and when the number of variables gets close to the number of laboratories, the selection of variables becomes less meaningful. Thus with an awareness of the sensitivity and fragility of the methodology, let us proceed to an examination of the variables selected by the two different processes.

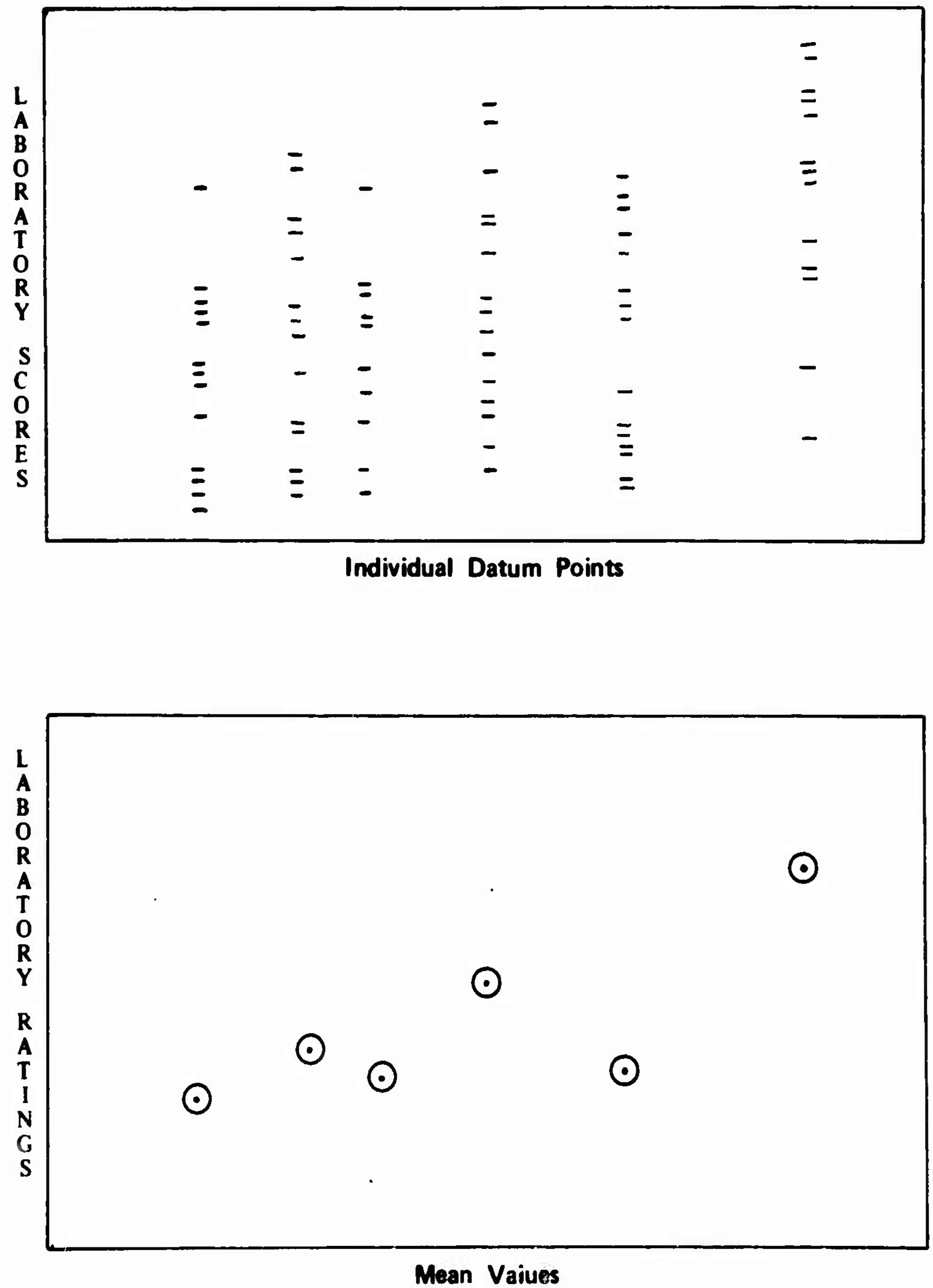


FIGURE 8.1

Diagram of Individual vs Mean Values

8.3 Multiple Regression

The Dahlgren Multiple Regression and Comprehensive Analysis Model was used to determine what, if any, relationship exists between the rankings or ratings and the seventy quantitative laboratory properties in the fiscal year 1969 data base. Letting n_j be the number of laboratories ranked by rater j, and v_{ij} be the rank of the i^{th} laboratory ranked by rater j, the numerical value of the rating or reputation of laboratory i by rater or judge j was taken as

$$r_{ij} = \frac{n_j - v_{ij}}{n_j - 1}$$

This placed all ratings on a (0,1) scale, 0 meaning low reputation and 1 meaning high. For example, suppose rater five ranked laboratory eight fourth out of a total of forty laboratories which he ranked. Then

$$r_{85} = \frac{40 - 4}{39} = \frac{36}{39} = .923$$

This provides a numerical value of reputation or rating for each laboratory by each rater who rated the laboratory. It is well to note that, in general, each rater rated a different number of laboratories.

Several models were postulated and will be discussed below. However, the feature consistent with all the models used was the ratings (the r_{ij}'s) were considered as the dependent variable and the quantitative laboratory property or variables were considered as the independent variables. The main objective was to determine the set of independent variables which had the strongest relationship with the dependent variable. Not all seventy independent variables could be used due to program and theoretical limitations. The number of independent variables used ranged between forty-three and forty-six, those deleted from the set of seventy being those with the lowest simple correlation with the dependent variable. Also, not all DoD laboratories were considered. Only the non-medical laboratories (fifty-six) were considered and of these, five were deleted because they had properties similar to medical laboratories. In addition to this, not all ratings were used. Those ratings used were from raters who rated in excess of ten non-medical laboratories. This reduced the number of raters to 280 as shown in Figure 2.14.

Hence, the analysis was conducted using forty-three to forty-six laboratory variables, fifty-one laboratories, and ratings from raters who rated in excess of ten non-medical DoD laboratories.

As a first attempt, it was decided to employ regression techniques to fit rating as a function of forty-five laboratory variables. The results are shown in Table 1 and 1A of Appendix M; the salient statistics, including an analysis of variance (anova) table, the coefficient of multiple determination, the coefficient of multiple correlation and the standard error are shown in Figure 8.2. In the anova table. the total sum of squares (SS) is simply the

$$\sum_{ij} (r_{ij} - \bar{r})^2$$

where $\bar{r}$ is the overall average rating. The sum of squares due to regression is the portion of this total which is explained by the regression equation. The ratio of the regression sum of squares to the total sum of squares is called the coefficient of multiple determination, and is denoted as

$$R^2 = \frac{\text{Sum of Squares Due to Regression}}{\text{Total Sum of Squares}}$$

The square root of R^2, or R, is the coefficient of multiple correlation; note that if R = .7, less than half of the total variation is due to or explained by the regression equation.

In the present case, the total sum of squares is 559.7200 and the sum of squares due to regression is 127.4787, whence R^2 = .2278. Hence, less than 23% of the total variation is due to or explained by the regression equation. The remaining 77% is the residual sum of squares, or the sum of squares about the regression equation. This quantity can be divided into two components. The first is the sum of the squares due to pure error, or the sum of the squares accounted for by repeat ratings of each laboratory by many raters. The remaining portion is the sum of the squares due to lack of fit (l.o.f), that is, due to the fact that with fifty-one laboratories, fifty laboratory variables could have been fit and only forty-five were actually fit. The coefficient of multiple correlation is simply the square root of the coefficient of multiple determination and is a measure of the association between rating and all the forty-five laboratory variables combined. Finally, the standard error of estimate is the square root of the error mean square, and represents how well the regression equation can be used for prediction.

Some amplification will now be given with respect to the standard error estimate. Consider a hypothetical laboratory whose laboratory variables have values equal to the averages of the fifty-one non-medical laboratories. Suppose now that one wished to use the regression equation to predict the rating of this laboratory. One would be about 95% confident that the rating by the single rater would lie within ±2 standard errors of the predicted rating. Obviously, this is poor precision since the ratings are on a [0,1] scale and the width of the confidence interval is about 1.06. Hence, one could not use the regression equation to predict, with any accuracy, the rating of an individual rater or judge. However, if one were content to predict the average rating of k raters, this could be done with a fair degree of precision. For example, assuming again that we are dealing with a hypothetical laboratory with average values for laboratory variables, one would be about 95% confident that the average rating from k raters would lie within ±2 s.e./$\sqrt{k}$ of the predicted rating. Should k = 100, this would be .5314/10 = .0532, yielding a confidence interval of about one decile. It should be re-emphasized that this predictive accuracy is based on a laboratory whose variables have average values over the fifty-one laboratories considered. For any realistic laboratory, where the variables differ from these average values, this accuracy will decrease.

The results shown in Figure 8.2 are based on fitting forty-five laboratory variables. The first variable selected is the one with the highest simple correlation with rating; this variable was the number of civilians with doctorate degrees (CIVPH). This single variable provided a correlation of .3031, a sum of squares due to regression of 51.45 (opposed to 127.52 for all forty-five variables) and a residual mean square of .0825. The next variable selected was the one with the highest partial correlation with rating or equivalently, the highest additional sum of squares due to regression after fitting the first. This variable was NONMS which *along with* CIVPH provided a multiple correlation of .3500, sum of squares due to regression of 68.55 and a residual mean square of .0798. The sequence in which the variables were selected as well as the corresponding values of R, sum of squares due to regression, and the residual mean square are shown in Table 1A (the first fifteen values are shown in Figure 8.2). The values in the last row of Table 1A, corresponding to the last variable selected, corresponds to the values listed in Figure 8.2 and Table 1 for the multiple correlation coefficient, sum of squares due to regression, and the residual mean square.

The procedure for the order in which the variables were selected for inclusion in the model is, in the statistical community, referred to as the Forward Selection (or Step-Up) procedure. It provides a descending ordering of the independent variables with respect to their importance or association with the dependent variable. This ordering is shown as the first column in Table 1A and Figure 8.2. It is

CONFIGURATION

51 DOD Non-Medical Labs
45 Lab Variables
0 Block Variables

ANALYSIS OF VARIANCE

Source	DF	Sum of Squares	Mean Square
Regression	45	127.5200	2.8338
Residual	6118	432.2000	0.0706
LOF	5	.9000	0.1800
Pure Error	6113	431.3000	0.0706
Total	6163	559.7200	

Coefficient of multiple determination = .2278
Coefficient of multiple correlation = .4773
Standard error of estimate = s.e. = .2657

SEQUENTIAL STATISTICS

Lab Variable	Multiple Correlation	Cumulative Sum of Square	Residual Mean Square
CIVPH	.3031	51.45	.0825
NONMS	.3500	68.55	.0798
IHO&M	.3706	76.88	.0784
OH6.5	.3852	83.04	.0775
CFTGS	.3925	86.24	.0769
IH6.3	.4007	89.88	.0764
OTHRD	.4100	94.09	.0757
IH6.1	.4189	98.21	.0750
EA/PR	.4227	100.00	.0748
RPROP	.4311	104.04	.07[illegible]1
PAPER	.4348	105.82	.0738
OH6.2	.4383	107.54	.0736
SEQIP	.4445	110.59	.0731
LSPAC	.4483	112.47	.0728
IHMPE	.4528	114.77	.0724
OHMPE	.4556	116.16	.0722
HOUSE	.4601	118.49	.0718
OWNED	.4627	119.83	.0716
LEASD	.4649	120.95	.0715
CIVGS	.4656	121.34	.0714
OTHOM	.4666	121.85	.0713
MEETS	.4675	122.35	.0713
CIVBS	.4683	122.76	.0712
CLASS	.4693	123.27	.0711

FIGURE 8.2

interesting to note that the variables selected for regression are generally uncorrelated, as is shown in Figure 8.3. The first variable selected after CIVPH was NONMS; its correlation with CIVPH was .126. The next variable, IHO&M, has a correlation less than .13 with each of the two already selected; similarly the fourth and fifth elements have relatively little correlation with the variables already in the regression equation. Beyond this point, the newly selected variables begin to become more correlated with the elements already selected. Note that these first four additional elements have raised the correlation from .3034 to .3931 - an absolute increase of .0897 and a relative change of 33%. The next forty will only increase the correlation by 20%, with an absolute change of .0841.

	NONMS	IHO&M	OH6.5	CFTGS	IH6.3	OTHRD	IH6.1
CIVPH	.126	-.121	.039	.210	.428	.345	.836
NONMS		-.060	.167	.171	.162	-.036	-.006
IHO&M			.111	-.047	-.025	-.166	-.170
OH6.5				.002	.162	.466	.061
CFTGS					.306	.569	.083
IH6.3						.233	.276
OTHRD							.385

FIGURE 8.3
Correlations Between Variables Selected for Regression

The example discussed above did not include blocking variables; these were added to the remaining eight cases included in Appendix M, but at the expense of reducing the set of data elements to be considered for regression. The elements removed were OH6.1, OH6.2, CIVGS, and OWNED. Three blocking schemes were used to examine the variation between rater groups. Schemes I and II are shown below. Scheme III was identical to Scheme I except that the last group, Universities and Not-for-Profits, was combined with the DDR&E raters. The composition of the groups for Scheme II is the same as that described in Chapter 1. The composition of the groups in Scheme II is the same as was shown in Figure 2.15; the "Service Not Identified" Group includes the raters whose service identification was inferred in Chapter 2. Another difference between Schemes II and III was in the composition of the laboratory groups. In Scheme II, the Behavorial Sciences Laboratory was included with the Army non-medical laboratories, but the Aerospace Research Laboratory was not; in Scheme III, the Behavioral Sciences Laboratory and the Aerospace Research Laboratory were both omitted. Scheme II also included the Air Force Human Factors Laboratory; this was omitted in Scheme III.

I, III*	II
OSD(DDR&E)	OSD(DDR&E)
Army	Headquarters Staffs
Navy	Service Commands
Air Force	Laboratories
Service Not Identified	Other Government (Non-DoD)
Other Government (Non-DoD)	Industry
Industry	Universities and Not-for-Profits
Universities and Not-for-Profits	

*III is like I, except the first and last groups are combined.

The significance of the blocking variables can be measured by the ratio of the Blocking Mean Square to the Residual Mean Square. A ratio in the vicinity of "2" indicates that there may be significant biases between the rater groups delimited by the blocking variables; a ratio greater than three or four is very definite indication of rater group differences.

The results are summarized in Figure 8.4 ; more detail is provided in Appendix M. For each regression configuration, the first two lines tell the number of laboratories, the military department, the arrangement of blocking variables, and the ratio of the blocking mean square to the residual mean square. The next five lines containing the names of the first five data elements selected for regression. The first of the last two lines gives the multiple correlation coefficient using the first five elements: the second line is the multiple correlation coefficient obtained for all elements used. (For the individual military departments, the total number of elements used was always one less than the number of laboratories.) The ± signs preceding the data elements indicate whether they entered the regression equation positively or negatively.

There was little difference between the regressions for the fifty-one laboratories as a group: the correlations came out about the same, there was no indication of rater-group biases, and the first eleven variables were identical. There were considerably more differences in the regression characteristics at the service laboratories level. The Army and Navy laboratories were found to be subject to significant rater-group biases (see Section 2.4), and only the first two or three elements were identical within the different military departments. A large part of this latter effect is attributable to the change in the number of laboratories; it is possible that the change in the Navy selections was caused by changes that were made to the fiscal year 1969 data base between the computations using Scheme II and those using Scheme III.

As indicated earlier, the modus operandi in these various experiments was not to over-control them at the outset, but rather to first give the regression system free rein to follow its own autonomous mathematical machinations, and then to become more selective in the list of candidate regression variables. Thus, while the immediate relevance of data elements such as land LEASD and land OWNED seems to the author to be of lesser significance than the composition of the laboratory's professional staff or the spectrum of its RDT&E activity, they (and other variables) were permitted to enter the regression equation with the expectation that on the second go-around the variables would be selectively screened.[1]

[1] Variables such as these may well be of secondary relevance, in that they reflect the size of the facility, land available for test and evaluation, geophysical environment, etc. However, they do not connote scientific productivity to the extent of other variables such as Civilians with Masters Degree, Papers Published, Patent Applications, Meetings Attended, etc. - all of which were in the second group of five variables common to the Army laboratories.

Table 1 51 DoD Laboratories	Table 4 23 Army Laboratories Scheme II; Ratio < 2	Table 5 22 Army Laboratories Scheme III; Ratio > 7
+CIVPH	+LEASD	+LEASD
+NONMS	+CIVPH	+CIVPH
−IHO&M	−TAMIL	+OWNED
+OH6.5	+OWNED	−TACIV
+CFTGS	−IHO+M	+IH6.1
.3925	.4071	.4417
.4773	.4448	.4701

Table 2 51 DoD Laboratories Scheme I; Ratio $\ll 1$	Table 6 17 Navy Laboratories Scheme II; Ratio > 2	Table 7 17 Navy Laboratories Scheme III; Ratio < 3
+CIVPH	+CIVMS	+CIVMS
+NONMS	+IH6.3	+IH6.3
−IHO&M	−MILPA	−MILPA
+OH6.5	−IHO&M	+CIVPH
+CFTGS	+CIVPH	−NONRD
.3931	.5424	.5523
.4772	.5552	.5613

Table 3 51 DoD Laboratories Scheme II; Ratio $\ll 1$	Table 8 11 Air Force Laboratories Scheme II; Ratio > 4	Table 9 10 Air Force Laboratories Scheme III; Ratio > 13
+CIVPH	+PAPER	+PAPER
+NONMS	+CFTGS	+CFTGS
−IHO&M	+MEETS	+MEETS
+OH6.5	−MILND	−IH6.3
+CFTGS	−PATNT	−LEASD
.3929	.3190	.3395
.4754	.3228	.3411

FIGURE 8.4
Summary of Results from DA-MRCA

8.4 Stepwise Regression

The stepwise multiple regression differs from the MRCA model in that at each step a test is made to determine whether one or more of the variables already in regression equation is no longer (mathematically) needed, and can therefore be deleted. The sensing is based on a test of significance called the F-test, which depends upon the number of variables already in the regression equation and also upon a significance level specified by the user. In the computations described here, the threshold for the F-test was set equal to "1". All computations were made using the fiscal year 1968 data and the mean-value of the rankings for each laboratory. The insidious implication of this is that it is possible to get a perfect fit between the data elements and the peer ratings with a relatively small number of points; for example, the multiple correlation coefficient for the Air Force laboratories can be driven to one with no more than ten variables in the regression equation - but this is a mathematical rather than a meaningful relationship, since any ten variables would accomplish the same result. Therefore, one must not attach undue emphasis to the higher correlations obtained using the mean-value ratings.

Four configurations of data were examined for each of the three military departments; these are shown in Appendix N and in Figure 8.5. The first configuration - called Set A - consisted of thirty-eight of the seventy basic elements of the data base. These represent four of the six major categories: Personnel, Appropriations, RDT&E Sub-Appropriations, and Training/Productivity. The two missing categories are Facilities and Source of Funds. As previously noted, it was intended to repeat the examination with emphasis on the reserve elements, but as before, this was not accomplished. Thus the regression equation was deprived of elements such as EQUIP, RPROP, SEQIP, etc., and DEPRD, DEPPR, DEPOM, etc., in making its selection of variables.

The second configuration - Set B - consisted of the thirty-five expanded variables; hence some aspect of the facilities and source of funds were included in the elements of total space (TSPAC), acquisition of scientific and engineering equipment (SEQAS), land owned or leased (ACRES), and source of funds from own department, other DoD, or non-Dod (TDEP$, TODOD, TNDOD).

The third configuration consisted of a number of ratio variables that had shown higher-than-average correlations between the peer ratings and the fiscal year 1968 data. Most of these are intuitively recognizable as meaningful ratios or proportions but the nature of the relationship in a few cases is somewhat abstruse. These consist of NONRD/RPRTS, CIVND/TR&D$, PAPER/TR&D$, MEETS/TR&D$, and SEQAS/TR&D$. The first one - non-DoD source of research dollars per report -

REGRESSION CØNFIGURATIØNS

SET A	SET B	SET C
MILBS	MPRØF	CIVMS/TPRØF
CIVBS	CPRØF	CIVPH/TPRØF
MILMS	TPRØF	IHR+D/TPRØF
CIVMS	TRACH	PAPER/TPRØF
MILPH	TMAST	PATNT/TPRØF
CIVPH	TPHDS	T6.1$/TPRØF
MILND	TAPER	SEQAS/TPRØF
CIVND	TSPAC	TØ+M$/TPRØF
SEQNP	TR+D$	TNDØD/TPRØF
SEQPR	TPRØ$	MEETS/TPRØF
IHR+D	TØ+M$	T6.1$/TR+D$
ØHR+D	TPGM$	SEQAS/TR+D$
IHPRØ	TIH$	CIVND/TR+D$
ØHPRØ	IH1-2	TØ+M$/TR+D$
IHØ+M	IH1-3	MEETS/TR+D$
ØHØ+M	IH1-4	ØHØ+M/TR+D$
MILCN	ØH1-2	PAPER/TR+D$
IH6.1	ØH1-3	TIH$/CIVSV
IH6.2	ØH1-4	PATNT/CIVBS
IH6.3	TDEP$	NØNRD/RPRTS
IH6.4	TØDØD	T6.1$/DEPRD
IH6.5	TNDØD	T6.4$/DEPRD
IH6.6	T6.1$	
IHMPE	T6.2$	
ØH6.1	T6.3$	
ØH6.2	T6.4$	
ØH6.3	T6.5$	
ØH6.4	T6.6$	
ØH6.5	T61-2	
ØH6.6	T61-3	
ØHMPE	T61-4	
PATNT	TØH$	
PAPER	ACRES	
RPRTS	SEQAS	
CIVGS	TØTND	
MILGS		
CFTGS		
MEETS		

FIGURE 8.5

may be related to the peer ratings to the extent that a laboratory is known outside of the Department of Defense based on reports of RDT&E. The items normalized by the total RDT&E program probably indicate the proportion of activity according to RDT&E category, i.e., a high proportion of civilians without degrees is representative of the 6.3 to 6.5 areas of RDT&E, while a high proportion of meetings attended or papers published would represent the 6.1 to 6.3 categories.

The fourth configuration consisted of the same elements as in Set A, except that logarithms of the elements were used. In order to compensate for some of the elements being zero, the transformation was actually of the form LOG (X + K) rather than LOG (X). In the correlations previously computed using logarithms, the value of K was set equal to 1. However, in the present case I had converted the money elements into units of millions of dollars instead of thousands. The addition of a 1 to the elements would therefore mean adding one person to the people elements, but one million dollars to the financial elements. In lieu of re-expressing the financial elements in thousands of dollars, I let K = .1. This probably did not substantially affect the correlations with the values chosen, but whether or not a significantly different set of independent variables would have been selected for some other choice of K is not known.

The results of using the four sets of variables in each of the military departments are shown in Tables 5-16 of Appendix N. The notation *VARIABLE* means that the variable was "removed" from the regression equation at that step. The variables selected in the first five steps of each configuration and the coefficient of multiple correlation at the end of the fifth step are summarized in Figure 8.6. A minus sign preceding a variable indicates that it entered the regression equation negatively. The asterisk on the Navy correlation of the basic variables is to draw attention to the fact that the value shown is the correlation at the end of five steps, but that at the seventh step CIVMS was removed, so that the effective correlation with five variables is .9812. Similarly with the Navy ratio variables, the effect of removing CIVPH/TPROF and IHR&D/TPROF at the eighth and ninth steps produced a correlation of .9125 with five variables. The effects of two similar such removals in the Air Force configurations are already included in the results shown in Figure

In most cases, the regressions have been carried out to the saturation point, i.e., the point where the number of elements in the regression equation is equal to the number of laboratories under consideration. Thus the value of the regression equation as a guide to laboratory management first tends to increase as more variables are entered at the beginning, but then peaks and tends to reflect the irregularities in the data.

BASIC	LOG BASIC	EXPANDED	RATIO
Army			
IH6.1	IH6.1	T6.1$	CIVMS/TPROF
-CIVND	-MEETS	-TOTND	SEQAS/TPROF
OHO+M	OHMPE	TODOD	TNDOD/TPROF
MILND	OH6.5	TPROS	-PAPER/TR+DS
-SEQPR	-CFTGS	TSPAC	OHO+M/TR+DS
.8617	.8542	.8537	.8494
Navy			
CIVMS	IH6.1	TMAST	IHR&D/TPROF
IH6.3	MILGS	SEQAS	-CIVND/TR+DS
-OH6.2	PATNT	---	SEQAS/TPROF
OHMPE	-CIVGS	---	-PAPER/TR+DS
IH6.6	CFTGS	---	CIVPH/TPROF
.9705*	.9511	.9220	.8948*
Air Force			
-OH6.4	SEQPR	-T6.4$	-T6.4$/DEPRD
CIVPH	-IHPRO	TPHDS	MEETS/TPROF
-CIVGS	-IH6.4	-TO&MS	-OHO&M/TR&DS
IH6.3	-PAPER	SEQAS	TNDOD/TPROF
CFTGS	MEETS	ACRES	PATNT/CIVBS
.9972	.9877	.9899	.9987

*Explanation in text.

FIGURE 8.6
Stepwise Regression, FY 68 Data

For example, in the regression using the basic configuration of variables (Set A) for the Army laboratories, the first element selected is in-house research appropriations (IH6.1), which had a simple correlation of .541 with the peer ratings. The next variable selected was non-degree civilian professionals (CIVND), which augmented IH6.1 to raise the correlation to .734 - even though the simple correlation between CIVND and the peer ratings was only .072. The addition of out-of-house Operations and Maintenance (OHO&M) at Step 3 raises the correlation to .785, and of non-degree military professionals at Step 4 to .845. Note that the relation between IH6.1, CIVND, and the peer ratings is such that CIVND is entered negatively, even though by itself it had a slightly positive correlation. On the other hand, MILND has a simple negative correlation, but is used positively in the regression equation. Similarly at Steps 5 and 6, SEQPR and IHPRO are added with signs just the opposite of their individual correlations with the peer ratings.

With respect to this point, Draper and Smith [13] point out that "it must be remembered that the model will be used by some people who are unaware of the fact that the least-squares regression coefficients are adjusted for other variables in the regression. Thus, they may attempt to predict the response by changing only one variable, using its coefficient to decide how much to change it. If all the estimated coefficients are independently estimated, this may do little harm. However, when the independent variables are highly correlated and the estimated coefficients are also correlated, reliance on individual coeffiients can be dangerous. It is wise to restrict prediction to the region of the X-space from which the original data were obtained, in any event. A check can also be made to see if individual coefficients are directionally correct, for example, if X_1 is the amount of production and Y is the total yield, then the coefficient b_1 should be positive."

At the end of six steps, the regression equation is of the form

$$\begin{aligned} R = A_0 &+ .00004 \times \text{IH6.1} \\ &- .00250 \times \text{CIVND} \\ &+ .00009 \times \text{OHO\&M} \\ &+ .01667 \times \text{MILND} \\ &+ .00003 \times \text{SEQPR} \\ &- .00002 \times \text{IHPRO} \end{aligned}$$

The coefficients of the elements with units in dollars are smaller than those for CIVND and MILND because the money elements are so much larger than the people elements - even though the money elements are expressed in units of thousands of dollars.[1] The coefficients continue to change throughout the regression, according to the variables as new ones are added. Thus the coefficients for the above elements at the end of the tenth step are

IH6.1:	.00007
CIVND:	-.00155
OHO&M:	.00010
MILND:	.00707
SEQPR:	.00003
IHPRO:	-.00002

The first variable to be removed is MILND, which is deleted at Step 12. But then it is re-entered at Step 17, this time negatively. In another turn-about, OH6.3, which was added at Step 14, is deleted at Step 19. From this behavior, I would judge that the regression should not be taken seriously beyond the twelfth step (at which point it contains ten variables); and on the basis of the change in R^2, probably only the first six steps should be considered.

The results of using the other configurations for the Army, Navy, and Air Force laboratories are shown on pages 5-10 of Appendix N. In two cases - TODOD in the Army expanded variables, and OH6.4 in the Navy logarithms of basic variables - the coefficients of the variables changed sign while they were in the regression equation; in both cases, they were deleted on the succeeding step. Only two elements - TMAST and SEQAS - were obtained from the Navy expanded set; this is because at Step 3, none of the remaining variables could cough up an F-ratio larger than $F_0 = 1$. In the configuration applying ratio variables to the Army laboratories, Steps 9 and 10 show an interesting example of apparently opposite effects; Step 9 enters T6.1\$/DEPRD positively, while Step 10 enters T6.1\$/TR&D\$ negatively. The former probably reflects the effect of the reputation of the laboratories in attracting funds from outside their own department; the latter is probably negative because of the vagaries of the regression process, for as noted in Section 4.3, the proportion of 6.1\$ to total R&D\$ was significantly positive for both the Army and Navy laboratories.

---

[1] The numerical values of the coefficients would have contained more significant figures if I had appropriately scaled the variables in advance.

Additional insight to the selection process may be obtained from Tables 2, 3, and 4 of Appendix N. These show the various basic variables that were candidates for selection at each of the first ten steps in Set A. The data for the first five Air Force steps are shown in Figure 8.7. The program first selected OH6.4, because that variable had the largest single correlation with the peer ratings. At Step 2, the program computed the partial correlation of the remaining variables to see which would contribute the most to the regression, subject to OH6.4 already having been entered into the regression equation. Any one of the five variables might have been a suitable choice at Step 2, but since IH6.2 had the highest partial correlation (-.69), it was chosen. At Step 3, the program again computes the partial correlations of each element with those already entered into the regression equation; CIVPH, IH6.1, and IH6.3 are each about equal candidates. At the next step, CIVGS is selected as the fourth variable to enter the regression equation.

While the entering of variables has been going on, at each step the procedure has also been examining the variables already entered to see if any can be deleted. Step 5 is such a deletion step; it is found (see Appendix N, Table 13) that IH6.2 can be removed with very little reduction in the coefficient of multiple determination.

The Air Force selections differ from the Army and Air Force processes, shown in Tables 2 and 3, in that two of the first five Air Force variables selected had been among the first five candidates in previous steps. Thus CIVPH, which was selected in Step 3, had been first to the top in Step 2; and IH6.3 - selected at Step 6 - had been third to the top in Step 3. In the Army selections, however, none of the other four candidates mentioned in Step 2 appears again in the first ten steps, except for MEETS, which is mentioned at Step 8. In fact, in five of the first six steps (excluding the first), the Army variable selected has not appeared among the first five in any of the previous steps, except for MILND, which was mentioned in Step 3. Similarly with the Navy variables, in the first five steps beyond the first, the variable selected has not appeared previously among the top five candidates. Navy Step 7 is the deletion of CIVMS; the variables selected first in Steps 8, 9, and 10 are ones that have been mentioned earlier.

These tables indicate the combinatorial sequences that could result from the regression analysis. At each step there are generally several eligible candidates; and since they are frequently inter-related, choosing one essentially eliminates the others from further consideration. Hence any small change in the selection process could quite radically alter the choice of subsequent variables.

AIR FØRCE - TYPE A

HIGH FIVE - EACH STEP

FIRST TEN STEPS

STEP 1	R	F
ØH6.4	-.83	17.3

STEP 2	R	F
IH6.2	-.69	6.5
CIVPH	.68	6.2
IH6.1	.65	5.2
ØHØ+M	-.63	4.6
IHPRØ	-.61	4.2

STEP 3	R	F
CIVPH	.57	3.0
IHR+D	.56	2.8
IH6.1	.54	2.4
IH6.3	.49	1.9
SEQNP	.48	1.8

STEP 4	R	F
CIVGS	-.63	3.3
ØHMPE	-.59	2.7
CIVBS	-.53	2.0
IHMPE	.48	1.5
ØHØ+M	-.46	1.3

STEP 5*	R	F
IH6.3	.91	20.0
IHMPE	.83	9.1
CIVMS	.79	6.6
CFTGS	.73	4.6
MILCN	.71	4.0

STEP 6	R	F
IH6.3	.91	24.5
IHMPE	.85	12.8
CIVMS	.81	9.5
MILCN	.74	5.9
CFTGS	.71	5.2

FIGURE 8.7

8.5 Regression Across Years

The preceding analyses were conducted for fiscal year 1968 only, since at the time they were initiated, not all of the three years of data was available.[1] In order to look at the relationships across different years of the data base, a simple least squares polynomial was used to fit the Navy and Air Force ratings to some of the data averaged over the three year period. The polynomial was then applied to the data for each year separately. The results for one of the Air Force data sets is shown in Figure 8.8. The fitted polynomial was of the form

$$\hat{S} = .0072 + 9.80 \left(\frac{T61 - 2}{TPGM\$} \right) + .0195\,(CIVPH)$$

S is the relative rating of the Air Force laboratories (normalized between 0 and 10), $\hat{S}$ is the polynomial approximation to S, E is the residual difference between S and $\hat{S}$, and ESQ is the sum of the squares of the residuals. For the average data, the coefficient of multiple determination, R^2 can be found from the relation

$$R^2 = 1 - \frac{\Sigma\, e^2}{\Sigma (S - \hat{S})^2} = 1 - \frac{22.6510}{76.0526} = .702$$

whence the coefficient of multiple correlation is found to be R = .838. The simple polynomial approximation is able to distinguish the two highest-rated laboratories and the lowest-rated one, but those in between are pretty muddled. I imagine this is largely a consequence of the original ratings being clearly bunched to begin with, making it difficult to discriminate between them.

A similar polynomial approximation to the Navy ratings was of the form

$$\hat{S} = 2.275 + .01237\,(CIVPH) + .7476\,(IHR\&D) - .3528\,(IHO\&M)$$

where the IHR&D and the IHO&M appropriations are in units of millions of dollars. The values of S and $\hat{S}$ for the different years are shown in Figure 8.8. The three highest-ranked and the lowest-ranked are consistently recognized as such across the different years; and except the twelfth-ranked laboratory, the $\hat{S}$ of each of the first eight are greater than the $\hat{S}$ of any of the last nine.

---

[1] It had been necessary to re-punch the FY 67 data.

RESIDUALS OF LOW-ORDER REGRESSIONS FOR DIFFERFNT FISCAL YEARS

SCORE	AVERAGE		FY 67		FY 68		FY 69	
S	$\hat{S}$	E	$\hat{S}$	E	$\hat{S}$	E	$\hat{S}$	E
AIR FORCE LABORATORIES								
10.00	10.12	-.12	9.69	..31	10.57	-.57	10.20	-.20
8.63	10.18	-1.55	10.16	-1.53	10.16	-1.53	10.23	-1.60
7.87	7.07	.80	6.71	1.16	7.97	-.10	6.82	1.06
7.83	5.04	2.79	4.73	3.11	5.04	2.79	5.26	2.57
7.59	6.07	1.52	4.86	2.73	5.83	1.76	7.90	-.31
6.95	5.57	1.38	4.81	2.14	5.24	1.71	8.01	-1.06
5.66	6.07	-.41	5.87	-.21	6.61	-.95	5.85	-.18
5.66	6.34	-.67	5.24	.42	9.84	-4.17	4.90	.77
3.21	5.05	-1.84	5.07	-1.86	5.07	-1.86	5.02	-1.81
.00	1.90	-1.90	2.09	-2.09	2.04	-2.04	1.68	-1.68
	$\Sigma\epsilon^2 = 22.65$		$\Sigma\epsilon^2 = 33.55$		$\Sigma\epsilon^2 = 42.44$		$\Sigma\epsilon^2 = 18.24$	

NAVY LABORATORIES

S	$\hat{S}$	E	$\hat{S}$	E	$\hat{S}$	E	$\hat{S}$	E
10.00	10.81	-.81	11.60	-1.60	10.68	-.68	10.15	-.15
8.41	7.54	.87	7.34	1.07	7.91	.49	7.33	1.07
7.56	6.26	1.30	6.54	1.02	6.28	1.28	5.97	1.59
5.95	3.79	2.15	4.47	1.48	3.63	2.32	3.27	2.67
5.42	3.50	1.92	3.32	2.10	3.53	1.89	3.64	1.78
5.38	3.73	1.66	3.77	1.62	3.63	1.76	3.81	1.57
5.25	4.18	1.08	3.54	1.71	4.33	.93	4.67	.58
3.40	5.62	-2.22	5.22	-1.82	6.13	-2.73	5.51	-2.11
2.93	3.22	-.29	2.14	.79	2.51	.42	2.47	.46
2.44	2.63	-.19	2.45	-.01	2.72	-.28	2.76	-.32
2.03	2.88	-.86	2.86	-.84	2.81	-.78	2.99	-.96
1.95	3.44	-1.49	3.26	-1.31	3.57	-1.62	3.49	-1.54
1.84	1.43	.41	.34	1.50	1.91	-.07	2.02	-.19
1.25	2.91	-1.65	2.79	-1.53	2.96	-1.70	3.00	-1.75
1.18	2.41	-1.23	1.47	-.29	2.86	-1.68	2.91	-1.73
.24	1.64	-1.39	.89	-.65	1.99	-1.75	2.03	-1.79
0.00	-.74	.74	-2.94	2.94	-.19	.19	.89	-.89
	$\Sigma\epsilon^2 = 30.26$		$\Sigma\epsilon^2 = 36.91$		$\Sigma\epsilon^2 = 35.04$		$\Sigma\epsilon^2 = 34.96$	

FIGURE 8.8

8.6 Summary

The utility of the regression model in identifying elements which relate to laboratory quality and technical competence depends upon several assumptions: (1) the peer rating of a laboratory is a reliable measure of the quality of its productivity; (2) the model gives a sufficiently accurate representation of the peer ratings; (3) the laboratory properties are optimally selected; (4) the relationship is applicable over a span of time (either forwards or backwards, but preferably both); and (5) the relationship is based upon some underlying meaningful phenomena.

The reliability of the peer ratings was discussed in Chapter 2, from the point of view of consistency and stability. While the individual scores of the various rankers cover almost the complete gamut of deciles, the rating of the laboratory - based on the mean value of all the scores - lies within a one-decile interval of the statistically "true" score. The ratings were also seen to be dependent upon the background of the raters - whether from DDR&E, Army, Navy, Air Force, or Industry - but again, the relative rankings of the laboratories within a military department were fairly consistent from rater group to rater group (except for the Industry ratings). The question at hand concerns a different aspect of reliability: the extent to which the ratings measure technical competence.

The rating of a laboratory is based on a number of different factors: the military department to which it belongs, the breadth and scope of its R&D program, the extent to which it has publicized its work, etc., *and* its technical competence. My impression is that the rating reflects the R&D capability of a laboratory more than its overall competence "to accomplish its assigned mission". One might ask: if the mission of a laboratory is not to perform R&D, what is it? This would lead into a discussion beyond the author's capability, and one to which consideration has been given by other writers. I would simply say that in the performance of R&D, the laboratories develop a certain technological expertise, which is sometimes diverted to the solution of immediate and important problems that do not in themselves involve further R&D, but rather represent the application of existing technology or methodology. It is oftentimes tempting, it is sometimes essential, for the laboratory to thus apply the fruits of its research to recurrent operational problems of an urgent or emergency nature - but there is also a risk of becoming overly involved with maintenance or engineering production, rather than going on to anticipate and provide for future problems. While I do not wish to dwell on this point, I suspect that some of the lower-rated laboratories may be doing a technically competent job with respect to their mission, but their mission may include tasks which are more operations and maintenance than research and development in nature.

Accuracy of the Model

How well the model represents the ratings is a function of how well the individual elements are correlated with the ratings, how they are correlated together, and how many are used. In the case of the Air Force laboratories, for example, it would be possible to develop an exact representation of the ratings by using ten of the laboratory properties; but this is simply a mathematical fluke - any ten properties could be used at random, in which case the relationship would generally not be meaningful. For most of the regression equations shown in the preceding sections, I would expect that at most ten terms of the equation would be the maximum to be used; and in most cases one should not go much beyond half the number of laboratories (i.e., for the Air Force laboratories, just use the first five terms of the regression equation).

Selection of Variables

The MRCA and the BMD02R regression models, having entered n variables into the regression equation, select as the next variable the one which will contribute "most" to reducing the variance between the observations (the ratings) and the data. This is not always a variable that is meaningful to the user: thus he has a problem of choosing meaningful variables to begin with, and thereby perhaps not reducing the variance as much as possible, or of using all variables, but having some which he does not quite understand. Part of the problem is that most of the *"meaningful"* variables - people with advanced degrees, equipment, R&D appropriations, etc. - are highly intercorrelated. If these are used, as in Section 8.5, the resulting regression equation is not as *"powerful"* as one with less intercorrelated variables. It also depends on how one intends to use the regression equation - for control, prediction, or just general information. In a general common-sense way, most of the laboratories already use the regression variables to *"control"* their technical competence and their R&D program. They are aware that a certain proportion of professionals with advanced degrees, or a certain proportion of research and exploratory development work, are essential to laboratory health and well-being (although they are not sure just what the proportions are).

Regression equations of the type shown could conceivably be used for predictive purposes, for laboratory managers to see "how things are going"; but I doubt that they will be used formally in this respect, (a) because of a natural skepticism about the automatism of such procedures and (b) because the skepticism is not unfounded. It is not entirely clear, even yet, what the ratings are measuring; and even if it were, it is not obvious that it has a future applicability (although the general agreement with the Apstein ratings, and the generally slow pace at which

organizations evolve, indicate that it would). It would be necessary to conduct a similar peer rating survey to more definitely determine the relationship with the laboratory properties and to develop a reliable predictive capability.

CHAPTER 9 CONCLUSIONS AND RECOMMENDATIONS

9.1 COMMENTS

(1) Other Rating Schemes

The ratings, as computed, assume that the rankings are linearly distributed from the highest "above average" to the lowest "below average". An alternative procedure would be to compute the ratings by the method of paired comparisons, as described by David [10] and Morrissey [11], but this has not been investigated. The method essentially counts the number of times one laboratory is preferred to another; this would be especially applicable to the Air Force laboratories because of the density of their ratings (many of the participants rated all ten Air Force physical sciences and engineering laboratories).

(2) Exclusion of Medical and Personnel Laboratories

The medical and personnel laboratories were excluded from the study because of their smaller size and because they were not as well known as the physical sciences and engineering laboratories. However, in several instances the correlations between the variables and the medical laboratories were substantially (but not necessarily significantly) higher than the corresponding correlations with the non-medical laboratories. This might be a topic for future investigation.

(3) Ratings Are Relative

It should be noted that the laboratory ratings are relative rather than absolute evaluations. To say a laboratory is "below average" is not to infer that it is of low quality, any more than to be "above average" implies being of high quality. To say that A is better than B establishes a relationship between them, but does not say anything about how good A is, or how poor B is, or even that they are good or poor. It depends on perspective: the optimist thinks this is the best of all possible worlds; the pessimist thinks so too.

(4) Qualitative Factors

It seems obvious that the technical competence of a laboratory depends more upon the quality of its leadership, the vitality of its mission, and the capability and enthusiasm of its people, than upon the number of people or the size of its technical program. However, the numbers of people and the amounts of dollars, plant, and equipment are the laboratories' basic resources; and in the

aggregate, the professional quality of the staff and the nature of the technical program are reflected by the proportion of personnel with advanced degrees and the proportion of funding for research and development. However, these are still only the resources; how they are applied and what is accomplished by them depends upon the organizational climate: encouragement of change, receptivity toward innovation, seeking out of meaningful and significant work, urgency of the program, etc.

9.2 SUMMARY

(1) Rankings per Rater

The average participant ranked twenty-three laboratories (including medical and non-medical). The average number of laboratories ranked per rater group varied from a high of twenty-seven (by the service headquarters group) to a low of twenty (by the industrial raters group); (this latter group had a large proportion of raters who ranked less than ten laboratories).

(2) Rankings per Laboratory

The medical and personnel laboratories received an average of thirty-eight rankings each. The physical sciences and engineering laboratories received an average of one hundred and twenty-eight rankings each.

(3) Computation of Ratings

The ratings were computed on a scale from 0 to 10, and ranged in value from 2.5 to 8.3. The standard deviations varied from 2.2 to 3.2; the 95% confidence intervals were between 0.6 and 1.6. The distribution of the ratings is considerably skewed; the distribution of the logarithms of the ratings is more centralized.

(4) Variation With Threshold

A number of experiments were conducted to test the sensitivity of the ratings under the requirement that each participant rank a minimum number of laboratories. The resulting rank-orders indicated that there were four major groupings of the laboratories: the first five were fairly invariant with respect to the size of the threshold (which ranged up to twenty); the next fifteen formed a group, but the order varied considerably within the group, according to the size of the threshold; the next nineteen formed a group similar to the second; and the last twelve formed a group similar to the first.

(5) **Consistency of the Ratings**

There is a wide variation in the ratings among the various rater groups. There is somewhat less variation among the various DoD groups than between the DoD groups and the industry group. However, there is general consistency about the high- and low-rated laboratories. The highest-ranked laboratories are mentioned among the higher-ranked laboratories of each major rater-group, and the lowest-ranked laboratories are mentioned among the lower-ranked laboratories of each major rater-group.

(6) **Variation by Rater Group**

There is a tendency for the DoD raters to rank the Navy laboratories higher than the Army laboratories (although this may be attributable to a few special cases, rather than being generically true); similarly, the industry group tends to rank the Air Force laboratories higher than those of the Army.

(7) **Dependency on Service Affiliation**

The service-affiliated raters (i.e., those in headquarters groups, service commands, and laboratories) tend to rank proportionally more of the laboratories in their own service than those in the others. The "average" Army rater ranked three-quarters of the twenty-three Army laboratories versus one-third of those of each of the other services; the average Navy rater ranked more than four-fifths of the eighteen Navy laboratories, but less than one-fifth of those of each of the other two departments; and the typical Air Force rater ranked all ten Air Force laboratories and three-tenths each of the Army and Navy laboratories. In several cases, there are strong indications of a preference to rate the laboratories in one's own department higher than those of the other departments; and in some instances there is more than a suspicion that laboratory raters have placed proprietary pride above objectivity.

(8) **Differences in Funding**

The technical program in the Air Force laboratories is funded almost entirely from R&D appropriations, whereas a much larger proportion of the funding in the Navy laboratories depends upon procurement, operations and maintenance, and miscellaneous appropriations. (The Air Force laboratories use a single program element - 6.1 for the research laboratories, 6.2 for the exploratory development laboratories - to cover the salaries of laboratory personnel, equipment acquisition, and related expenses, regardless of the category of work.) The corresponding proporitons in the Army laboratories lie between those of the other two services.

(9) Stability of Elements

The personnel elements of the quantitative laboratory properties are fairly stable over the three years of the data base (fiscal year 1967, 1968, and 1969), most of the major elements having less than a twenty-five percent maximum annual change. The principal funding element - the RDT&E appropriation - varies somewhat more, but in most cases the maximum annual variation is less than fifty percent. The other sources of funding - procurement, operations and maintenance, and miscellaneous - generally have maximum annual changes greater than fifty percent.

(10) Correlations Between Elements

As might be expected, there is considerable correlation between the various elements of the data base. Altogether, 117 of a possible 2278 pairs of elements had a joint correlation greater than 0.7, for the most part these represent the principal combinations of the quantitative laboratory properties.

(11) Distribution of Elements

The distributions of the data base elements among the various laboratories are characterized by their asymmetries. A few laboratories tend to have considerably more of a property than most of the others. For each data element, less than ten percent of the laboratories account for more than twenty-five percent of the value of the element.

(12) Variation With Fiscal Year

The correlations between the peer ratings and the laboratory properties are in many cases quite similar for each of the three years of the data base, even though some of the properties have fairly large annual variations. The most significant correlations are between the peer ratings and the properties of the Navy laboratories. The Army laboratories for the most part show only minor correlations between the peer ratings and the data base elements, although some experiments with a subset of the Army laboratories indicate that in selected circumstances the correlations may be comparable to those found for the Navy variables. A few of the Air Force correlations are as large as those found for the Navy, but because of the much smaller number of Air Force laboratories, most of the correlations are not statistically significant.

(13) Variation of Correlations by Rater Groups

Denoting by "DSC" the ratings obtained by pooling the rankings of the participants from DDR&E, Headquarters Staffs, and Service Commands, the

correlations between the peer ratings and the data base elements in Army laboratories tended to be highest for the DSC ratings and lowest for the ratings based on the rankings of the laboratory group. In general, the Army correlations were higher for the DoD ratings than for the Industry ratings. Among the Navy laboratories, there was not much difference in the correlations among the various DoD groups, but the DoD correlations were substantially higher than those of the Industry group. In the Air Force laboratories, the laboratory correlations were higher than those using the DSC ratings; and the correlations based on the DoD ratings were strikingly different from those of the Industry group.

(14) Dependencies on Extreme Points

Various examinations were conducted to examine the dependency of the correlations upon extreme points, or upon the largest or highest rated laboratories in each military department. Overall, these various examinations of the possible effects of outliers indicate that, at least for the Army and the Navy, there are relatively few cases where an extreme point unduly (1) raises a correlation to a significantly high value, or (2) masks out significant correlations in the remaining variables. The marginal number of Air Force laboratories precludes making a similar statement, one way or the other, about the effect of extrema on their correlations. A few of the higher Army correlations were dependent upon the highest rated laboratory; for example, the correlation between total procurement appropriations changed from .346 to -.304 with the deletion of the highest rated laboratory; the correlation with non-DoD source of funding dropped from .569 to .280. The Navy laboratories showed some dependency on both the highest-ranked and the largest laboratories, but for the most part, the correlations held up fairly well. The Air Force correlations were generally higher without the largest laboratory, and lower without the highest-ranked laboratory.

(15) Ratio Variables

In one experiment, all possible combinations of ratios of variables were compared with the peer rating. For the most part, these ratio correlations are about the same or less than those obtained from the original variables. In some cases, the ratio variables tend to bring the differences between the regular variables into sharper focus. Thus the correlations for professionals with no degree, professionals with bachelors degree, professionals with masters degree, and professionals with doctoral degrees are as follows:

	Army	Navy	Air Force
TOTND	.002	.150	-.102
TBACH	.218	.683	-.452
TMAST	.362	.881	.416
TPHDS	.368	.799	.621

but when normalized by dividing them by the total number of professionals, the variables have the following correlations:

	Army	Navy	Air Force
TOTND/TPROF	-.381	-.361	.070
TBACH/TPROF	-.266	-.174	-.668
TMAST/TPROF	.474	.115	.654
TPHDS/TPROF	.205	.421	.583

(16) Navy Correlations

With respect to the seventeen Navy laboratories, the higher-rated laboratories are those with the larger amount of certain key variables, i.e., the relationship depends on the size of the variable. Based on the correlations between the peer ratings and the Fiscal Year 1967-68-69 Average Data, the key Navy variables are the number of People with Advanced Degrees, Value of Equipment, In-House Research and Development (total of all categories), Military Construction, and Acquisition of Scientific Equipment. Each of these has a correlation greater than .800 with the peer ratings.

These same variables, or ones very similar to them, also serve as lines of demarcation between the higher-rated Navy laboratories and the lower-rated ones. For example, the first eight Navy laboratories each had a larger In-House RDT&E Appropriation than did any of the last ten; the first nine Navy laboratories each had more Professionals than did any of the last nine; and the first ten Navy laboratories each reported a higher value for Equipment than did any of the last eight.

The Navy correlations are generally smaller when they are computed using the ratio variables. For example, of the variables normalized by the total number of professionals, the only element with a correlation greater than .700 is IHR&D/TPROF, which has a correlation of .764.

(17) Army and Air Force Correlations

With respect to the twenty-three Army laboratories and the ten Air Force laboratories, not as much can be said for the unnormalized variables. The only Army correlations greater than .400 are In-House Research dollars (IH6.1; .524), total funding from sources other than DoD (TNDOD; .468), and land owned or leased (ACRES; .413). The only Air Force correlations greater than .600 are the total number of PhD's (TPHDS; .630), the in-house advanced development program (IH6.3; -.622), the engineering development program (T6.4$; -.857), and the number of papers published (PAPER; .598). From these, it might be inferred that the ratings are dependent on how well the laboratory is known, since the ability to attract funds from other departments would seem to be based on reputation and competence; similarly, it would seem that competence and reputation would be judged by the success of the research program, much of which is conducted by professionals at the doctorate level, who publish their findings in the technical literature.

Both the Army and the Air Force laboratories show a larger number of negative correlations when the elements of the data-base are normalized by the number of professionals. In both the Army and the Air Force laboratories, the highest (positive) correlations are the proportion of civilian professionals with masters degrees. Other correlations greater than .400 for the Army normalized variables are OH6.1/TPROF (.432), OH6.5/TPROF (.445), and IH1-2/TPROF (.425). For the Air Force, the normalized variables greater than .600 are TBACH/TPROF (-.669) and T6.4$/TPROF (-.793); other ratios greater than .500 are CIVPH/TPROF (.591), IH6.1/TPROF (.569), PAPER/TPROF (.531), and SFQAS/TPROF (.504).

(18) High-Low Comparisons

The two previous summaries were made for correlations utilizing all or most of the laboratories of the individual military departments. Studies were also conducted using only the few top-rated and the few bottom-rated laboratories in each service and in DoD as a whole. While quantity is still a principal factor in discriminating between the two sets - the top-rated laboratories have twice as much or more of the properties positively correlated with the peer ratings - there are also marked differences in the proportion of elements between the high groups and the low groups. The elements having the greatest differentials are professionals with advanced degrees, research dollars, amount of funding from outside the Department of Defense, acquisition of scientific equipment, and acres of land.

The higher-ranked laboratories also tend to have proportionally more of the key variables per professional than do the middle-ranked laboratories or the lower-ranked laboratories. Thus the higher-ranked laboratories consistently have a higher proportion of Professionals with Advanced Degrees, as well as a higher proportion of Research Dollars per Professional (both In-House and Out-of-House).

9.3 CONCLUSIONS

(1) The Bias in Peer Ratings Does Not Invalidate Them

The peer ratings are based on rankings that are subject to bias among the various rater groups, but the ratings have not been overly affected by these biases. This is because most of the bias is of a service-affiliated nature, and the principal portion of the rankings upon which the rating of a laboratory is based are rankings from the laboratory's own service. Hence there is a general agreement between the overall rank-ordering and the service rank-ordering of the laboratories within a particular military department.

(2) The Peer Ratings Are Generally Statistically Reliable

There is a great deal of variation in opinion among the individual raters about the relative quality of the various laboratories. There is also considerable variation of opinion between groups of raters, e.g., the DDR&E group, the Headquarters and Service Staffs, the Laboratory group, the Industry group, etc. However, when these are blended together, their size alone gives increased statistical validity to the ratings obtained from their union. On a scale from 0 to 10, the individual ratings generally lie within a 95% confidence interval of ±0.5 about their mean. Further, their general agreement with similar ratings produced by the Apstein survey in 1963 lends additional confidence concerning their statistical reliability.

(3) The Peer Ratings Are Meaningfully Related to the Quality of the Navy Laboratories

For the Navy laboratories, there are substantial correlations (of the order .800 and higher) between the peer ratings and the laboratory elements such as Professionals with Advanced Degrees, Equipment, Scientific Equipment Acquisition, and the In-House RDT&E Program. The correlations are based more on the size of the particular properties than on their generic proportions. The particular properties are so intimately associated with R&D capability that it must be concluded that the ratings are meaningfully related to the R&D competence of the Navy laboratories.

(4) The Relationships With the Army and Air Force Laboratories Are Less Clear

For the Army and the Air Force laboratories, the association between peer ratings and quantitative properties is not clear. The Army laboratories consist of a large number of extreme variables; several of the higher-rated laboratories have quite different characteristics. Nine of the laboratories have less than two hundred professionals, which makes discrimination between them difficult. The correlations are somewhat higher when normalized by the number of professionals, but even then they are only marginally useful. The Air Force laboratories, on the other hand, are more similar in the distribution of their characteristics, but their ratings are also more closely bunched than those of the other two services, tending to be more in the upper middle part of the distribution of the ratings. Also, the relatively small number of laboratories reduces the statistical significance of correlations which are of the same magnitude of the other two services, and which are therefore *seemingly* as meaningful.

(5) The Raters Put Emphasis on Research Capability

There is some suggestion, when considering the few top-ranked and bottom-ranked laboratories on an overall DoD basis, that the raters may have put a premium on the research aspects of laboratory activity. Thus, two of the four most dominant expanded variables in the tables of the percent accounted for by the top and bottom DoD groups are the number of PhD's and the magnitude of the research program. (These are the variables labeled TPHDS and T6.1$; the other two most dominant variables are TNDOD and ACRES. T6.1$ also emerges as the most dominant variables in the top-top, bottom-bottom tables shown in Figure 7.8; TNDOD is a close second.) It is also possible that the observed phenomenum is a secondary rather than a primary effect, i.e., the laboratories having the larger research appropriations may also be the most widely known, and are being cited by renown (this is not to argue that the quality of the research program was not initially responsible for the renown).

(6) Which Year to Compare?

The reputations of the laboratories change slowly with time; all but three of the twenty-nine laboratories for which there were corresponding ratings from the Apstein 1963 survey were in remarkably good agreement with the ratings of the present survey. This spot-lights one of the unanswered questions of the present study: which of the years of the data base typifies the raters' knowledge of the laboratories? The ratings were coincident with fiscal year 1969; but even assuming

that the raters had current knowledge of the laboratories' technical competence, is the state of that competence dependent upon current values of the resources, or does it reflect the resources that were available two, five, or ten years earlier? More realistically, the raters may have had current knowledge of only a portion of the laboratories they rated, making even more uncertain the lag between resource potential and laboratory accomplishment.

(7) Control Variables

The regression analyses, using the MRCA results, show that a linear regression equation can account for about 30% of the variation in the Navy ratings, about 20% of the variation in the Army ratings, and about 12% of the variation in the Air Force ratings. Alternatively, using the mean values of the rankings, disregarding their statistical variablility, one can account for 90% of the variations in the separate services with eight Army variables, three Navy variables, and three Air Force variables. However, these can be selected in a variety of ways, and will generally not be independent of one another, so that the use of the regression equation to "control" the quality of the laboratories is quite unlikely. Yet in a very real sense, the candidate variables are all representative of control variables, for they are measures of the basic laboratory resources. Hence the input, and consequently the output - and correspondingly, the peer rating at some future date - is in some way a function of these basic elements.

(8) Prediction Variables

A linear regression equation might conceivably be used for predicting the future ratings of the laboratories, but it would be necessary to conduct one or two more surveys, for calibration and validation, before one could hope to arrive at a meaningful regression equation. Further, it would be necessary to determine a function to represent the time lag between a rater's estimation of a laboratory's technical quality and its actual present capability. As to the value of such a model if it existed, it might be a useful management tool for answering "what if" types of questions; but on the whole, I doubt that it would have practical utility.

9.4 Recommendations

(1) Make Technical Competence Better Known

The laboratories should enhance their reputations by making their products or contributions better known, especially across military departments. One way to

do this might be to prepare attractive and informative exhibits which might be displayed in the Pentagon concourse, at regional meetings, loaned to other activities, etc.

(2) **Examine the Ranking Process**

It would be generally worthwile to further examine the ranking process. How does one rank a set of laboratories? To what extent does prejudice or self-interest enter the picture? How does a single favorable or unfavorable experience with one person at a laboratory effect its overall reputation? Are there associations between laboratories, so that the rating of one is coupled with the rating of another? In this connection, it is interesting to note that of the thirty-nine participants who ranked exactly one medical laboratory, the highest-rated medical laboratory received the most number of votes; but among the raters who ranked exactly two medical laboratories, the one mentioned most often had been mentioned only once by the participants who rated exactly one medical laboratory. Whether this is simply a coincidence, or whether there was something about the laboratories that prompted the raters to consider this laboratory paired with others, is not known.

(3) **Repeat Survey**

A peer ranking survey similar to the one described herein should be conducted within a three to five year period of the 1969 survey. The participants should identify themselves as before, and additionally according to service affiliation. Care should be taken to ensure that the different rater groups are given appropriate representation. The raters should also indicate the boundary points of the various groups into which they initially assigned the laboratories (Above Average, etc.). A follow-up interview should be conducted with a sub-sample of the participants in order to obtain insight to the various alternatives that were considered during the ranking process.

(4) **Maintain Data Base**

The laboratory resources data base should continue to be maintained and expanded according to the needs of its users.[1] With the addition of the data for

[1] Cognizance of the data base has been assigned to the Army Office of Laboratory Management. This Office is presently updating the data base to include the data from fiscal year 1971.

fiscal year 1971, the laboratory properties data base will span a period of five fiscal years, and should be able to provide comprehensive data upon which to base analysis and prediction of past, present, and future trends.

(5) Parallel Studies

Consideration might be given to conducting a similar sort of study using a parallel data base which contains information on individuals rather than institutions. This is the Salary Compensation Survey, which was initiated in fiscal year 1968, repeated in fiscal year 1969, and in odd years thereafter. In addition to salary, the data contains information on education, occupation, papers and publication, etc. Portions of this data are being used by Esbeck and Balwally at Case Western Reserve University in connection with the REFLEX Project.

(6) Parallel Evaluations

Surveys such as the peer rating survey, and studies such as the one described in this report, should be integrated with qualitative evaluations of laboratories, such as are conducted triennially within the Army laboratories, or such as are conducted semi-annually by the Naval Research Advisory Council.

(7) Alternative Approaches

It would be interesting to examine other ways of grouping the laboratories, to see what relationships might be found with the peer ratings. For example, as shown in Figure 9.2, Glass has computed the ratings of the laboratories arranged according to type of function.

* * * * * * * * * *

I would like to conclude this report with a quotation from some remarks by Dr. Finn J. Larsen, then Principal Deputy Director of Defense Research and Engineering, made in an address before the Aerospace and Science Technology Branch, Scientific Research Society of America (RESA), at Bolling Air Force Base on June 24, 1966. "It is a truism, but one one I think worth repeating, that the best way to have a dynamic laboratory is to give it a dynamic mission. Where in the range of research, exploratory development, advanced development, engineering

Functional Area	No. of Labs	No. of Ratings	% of All Ratings	Normalized Score	$\bar{X}$ Rating
Research	5	638	8.5	4378	67
Sea Warfare Systems	4	443	5.9	2601	59
Medical	22	830	11.1	4828	58
Electronics	5	748	10.0	4024	54
Ordnance	15	2295	30.7	12222	53
Aerospace	8	890	11.9	4686	53
Materials	4	439	5.9	2308	53
Chemical and Biological Research	5	595	8.0	2954	50
Engineering	4	328	4.4	1542	47
Behavioral Sciences	6	369	3.5	1371	37
Totals	78	7605	99.9	40647	54

FIGURE 9.2
Ratings of Laboratories According to Functional Areas

development and operational systems development should the missions of the in-house laboratories be? Not too much pure research, I'm afraid - no more than in a very good industrial R&D laboratory - and the research should be concerned with an environment of military importance. And certainly not at the other end of the scale - in the design of engineering prototypes for production, since, in order to be truly effective, the designer for production has to live next door to the factory.

What does this leave for the in-house labs? One of the most exciting jobs of all. Taking new ideas and concepts wherever they may come from and synthesizing them into future systems and components. Then, if and when the laboratory scientists' brain children do go into production, wisely monitoring production and helping the producers avoid the mistakes previously made in advanced development."

REFERENCES

1. Apstein, M. *"Effectiveness of Military Laboratories as a Function of Contract Activity"*, **IEEE Transactions on Engineering Management**, EM-12, No. 2, June 1965, pp. 44-50.

2. Glass, E. M. *"Evaluation of R&D Organizations"*, **Research/Development**, January 1970.

3. Glass, E. M. *"Methods of Evaluating R&D Organizations,"* paper presented at meeting of Military Operations Research Society (MORS), 19 June 1970. To be published in **IEEE Transactions on Engineering Management**, February 1972 issue.

4. U. S., Department of Defense, Office of the Director of Research and Engineering. **Department of Defense In-House RDT&E Activities**. Prepared by E. D. Anderson, MAR 69-4, 30 October 1969.

5. Memorandum from The Deputy Secretary of Defense, "Demonstration Project on Reconciliation of Workload, Funds and Manpower" dated 30 December 1969.

6. Roman, D. D. **Research and Development; The Economics and Administration of Technology**. New York, Appleton-Century-Crofts, 1968.

7. U. S. Department of Defense, Office of the Director of Defense Research and Engineering. **RDT&E Activity Data - Instructions for Using the Data Bank.** Prepared by James R. Locher, III, and Eugene G. Haberman, MAN 69-2, 1 May 1969.

8. Fisher, R. A., and Yates, F. **Statistical Tables For Biological, Agricultural, and Medical Research.** New York, Hafner Publishing Company, Inc., 1963.

9. Draper, N. R., and Smith, H. **Applied Regression Analysis.** New York, John Wiley and Sons, Inc., 1966.

10. U. S., Department of Defense. Office of the Director of Defense Research and Engineering. **Correlation and Regression Techniques.** Prepared by John W. McCloskey, MAN 67-1, October 1967.

11. David, H. A. **The Method of Paired Comparisons,** New York, Hafner Publishing Company, Inc., 1963.

12. Morrissey, J. H. *"New Method for the Assignment of Psychometric Scale Values from Incomplete Paired Comparisons"*, **Journal of the Optical Society of America**, Vol. 45, No. 5, pp. 373-378, May 1955.

Other References

Glass, E. M., *"Institutional Analysis of Compensation for Scientists and Engineers,"* **IEEE Transactions on Engineering Management**, Vol. EM-13, No. 1, March 1966.

R. W. Harrold, *"An Evaluation of Measurable Characteristics Within Military Laboratories,"* **IEEE Transactions on Engineering Management**, Vol. EM-16, No. 1, February 1969.

U. S., Department of the Navy, **RDT&E Management Guide**, NAVSO P-2457, 1 July 1969.

U. S., Department of Defense, **Report of the Task Group on Defense In-House Laboratories**, 1 July 1971.